Special Report 69

U.S. ARMY RESEARCH INSTITUTE
PROGRAM IN BASIC RESEARCH—FY 2010

Basic Research Unit
Paul A. Gade, Chief

November 2010

**U.S. Army Research Institute
for the Behavioral and Social Sciences**

**Department of the Army
Deputy Chief of Staff, G1**

Authorized and approved for distribution:

Michelle Sams

**MICHELLE SAMS
Director**

Research accomplished under contract
for the Department of the Army

Technical Review by

Paul A. Gade, U.S. Army Research Institute

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) November 2010	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) October 2009 to September 2010

4. TITLE AND SUBTITLE U.S. Army Research Institute Program in Basic Research—FY 2010	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER 611102

6. AUTHOR(S) Basic Research Unit	5c. PROJECT NUMBER B74F
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-BR 2511 Jefferson Davis Highway Arlington, VA 22202-3926	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, VA 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Special Report 69

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Subject Matter POC: Dr. Paul A. Gade

14. ABSTRACT *(Maximum 200 words)*:

This document contains detailed summaries for each of the U.S. Army Research Institute's basic research contracts for the fiscal year 2010. These summaries are grouped according to four Basic Research Unit program objectives: providing fundamental knowledge to improve training in complex environments; providing fundamental knowledge to improving leader and team performance; providing fundamental knowledge for identifying and measuring the attributes and skills that are critical to Soldier recruiting, assignment, performance, and retention in the transforming Army; and providing fundamental knowledge for organizational behavior and network science research. In addition to summarizing what was done or is being done, each summary describes the contributions of that research effort to basic behavioral science and suggests how the findings might benefit the Army and other military services.

15. SUBJECT TERMS
training, leadership, team performance, assessment, adaptability, motivation, social networks, work behavior, network science

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON Ellen Kinzer Technical Publication Specialist 703-602-8049
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	115	

U.S. ARMY RESEARCH INSTITUTE PROGRAM IN BASIC RESEARCH - FY 2010

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

THE BASIC RESEARCH PROGRAM: IDENTIFYING
NEW AND PROMISING TECHNOLOGIES

The Basic Research Unit (BRU) research program focuses on providing the basic research underpinnings for the personnel, leader development, and training requirements of the future. This program is a critical link between the world of behavioral science and the military community. Searching out and advancing the state-of-the-art methods, theories, and findings in behavioral science; encouraging projects most likely to contribute generalizable scientific principles and new knowledge; and supporting those efforts that have potential military relevance and likelihood of leading to applied behavioral technology are BRU's key research goals.

Universities conduct most of the basic research in the program. BRU maintains close contact with other basic behavioral science research organizations as well as with ARI's applied researchers, other relevant agencies within the Army, and other military services. These contacts help to define issues that require fundamental research and facilitate the transition of basic research results to applied programs for eventual use by the operational Army.

In BRU's contract programs, a Broad Agency Announcement (BAA) is issued each year to solicit both concept papers and formal proposals relating to the announced program research areas. In a given year, the BAA highlights the research objectives of special interest and provides an open call for proposals.

On the following pages, the reader will find summaries of current and recently completed BRU contracts, which began between 2007 and 2009. There are four current BRU program objectives, each of which is discussed in detail below. These objectives are:

1. To provide fundamental knowledge to improve training in complex environments;
2. To provide fundamental knowledge to improve leader and team performance;
3. To provide fundamental knowledge for identifying, assessing, and assigning quality personnel in the evolving Army; and,
4. To provide fundamental knowledge for understanding and evaluating organizational behavior and network science research.

Training in Complex Environments

Basic research in this area focuses on developing concepts and methods for training complex tasks and for sustaining complex task performance. This body of research also intends to assess the cognitive impact of technological requirements arising from digital, semi-automated, and robotic systems on training requirements. One of the efforts in this area evaluates an error-based approach in training complex interpersonal skills as opposed to the traditional behavioral modeling approach. Another effort is examining how affect influences what people see and their ability to detect and act on visual information in the absence of subjective awareness; this effort will contribute to training on perceiving threats. The models and theories produced under this objective should be useful in accounting for individual differences in training and facilitating practical, individualized, adaptive training methods. This body of

research will lead to advances that will translate into improved training methods in a wide range of tasks.

Leader and Team Performance

Commensurate with the requirements for rapidly developing adaptable, flexible leaders, the basic research program in improving leader and team performance is directed toward providing concepts and methods for accelerating leader development and understanding and developing leader adaptability and flexibility in a manner that can be tested in the applied environment. One effort is exploring the impact of social stereotypes on leadership performance. This project examines individual differences that potentially buffer or increase the susceptibility of certain groups to perform poorly as leaders, due to heightened anxiety of being judged or stereotyped. Another effort is focused on developing both adaptive leaders and teams through formal instruction, developmental work assignments, and self-development, while another project is identifying individual differences that are relevant in motivating leadership development. We are committed to discovering and testing the basic cognitive principles that underlie effective leader-team performance as well as understanding the dynamics of small group leadership in face-to-face and distributed team environments. Another of these team-centered efforts evaluates how people's language, specifically the use of function words, relates to group dynamics. Results from this body of research will make an important contribution to understanding and improving organizational effectiveness through improving the ways we develop our leaders and teams.

Quality Personnel for the Evolving Army

Identifying and measuring the aptitudes and skills that are unique to the human performance requirements of military service is a major theme of this basic research effort. As part of this process, we seek to devise methods that assess sociological and psychological factors that could influence recruitment, retention, and Army performance. One effort is developing new multi-media assessments of emotional abilities and linking these measures to outcomes such as decision-making, teamwork, and leadership behaviors. Another project is developing a knowledge test to assess personality and temperament such that respondents cannot improve their scores through faking. Another effort is concerned with developing a test of pattern recognition, as a predictor of mental flexibility, which could be applied in selection and placement; another investigation examines how perceived organizational support affects employee dedication and retention. Many of the basic research program's efforts fall under this objective; all of these research efforts vary while still contributing to research on the selection, placement, and retention of personnel.

Organizational Behavior and Network Science Research

Research under this objective is focused on using social networks, whether in simulations, games, or Army organizations to help understand and predict group behavioral processes. Modeling and understanding the networks underlying human knowledge systems is also part of this objective. One effort under this objective is studying what factors contribute to successful collaboration across distributed work groups. Another effort uses dynamic network

analysis (DNA) to examine how officers' organizational networks change over time. These and other efforts under this objective intend to study behavioral processes through the analysis of networks and changes within groups of individuals across a period of time.

This report provides a listing and brief synopsis of ongoing and recently completed research efforts. Project listings are organized into the four aforementioned research objectives. It is important to note, however, that basic research is but one of many programs for which BRU has responsibility.

Other programs in BRU include the following:

- The Small Business Innovative Research (SBIR) Program;
- The Small Business Technology Transfer (STTR) Program;
- The International Behavioral Science and Technology Watch;
- The graduate student apprenticeship program – the Consortium Research Fellows Program with the Consortium of Metropolitan Washington Universities;
- Outreach efforts to Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs), and research support in behavioral science for the U.S. Military Academy; and
- Multidisciplinary University Research Initiatives in training research and intercultural collaboration and negotiations.

Additional information about reports from these research efforts is available upon request.

Paul A. Gade, Chief
Basic Research Unit

Kresenda Keith, Consortium Fellow
Basic Research Unit

BRU RESEARCH OBJECTIVE #1: PROVIDE FUNDAMENTAL KNOWLEDGE TO IMPROVE TRAINING IN COMPLEX ENVIRONMENTS

Research under this objective develops concepts and methods for training complex tasks and sustaining complex task performance. The focus is on understanding the cognitive impact of Future Force technology on training requirements, the impact of voluminous multi-modal data on performance, and developing methods for improving cognitive skills.

A Research Plan to Understand the Impact of Training on Performance

Contract #: W91WAW-09-C-0081   **Contract Dates:** 9/29/2009 to 9/29/2014
Institution: Alion Science & Technology Corp.   **PI:** Sue Archer

Problem(s)/Research Question(s) – Do training methods significantly impact target task performance? Which combinations of training methods are more effective and efficient for training different types of tasks?

Technical Barrier(s) – Many factors interact to influence the effectiveness and efficiency of training for producing desired performance outcomes. While there is a broad and rich research literature on training effectiveness, the empirical evidence comparing the effectiveness of training strategies over different types of tasks is limited.

Significance/Impact for Basic Research – A better understanding of the effect of training methods and task factors on performance will add to the training effectiveness knowledge base, provide a foundation for empirically based training design guidelines, and improve predictive models of the effect of training factors on task performance.

Potential Transitions – The results of this research would have broad application in:
- ARI and other military and civilian applied training research programs.
- ARI and other military programs developing training design guidelines to help training managers and trainers identify the best ways to train specific task types.
- Human-system integration tools that explore the tradeoffs between training methods as they relate to retention, performance and transfer measures.

Overview

How does a training community make training decisions to maximize effectiveness and efficiency of training for complex cognitive tasks? Issues of cost effectiveness, training effectiveness, and training transfer are critical to the decision making process. There are many factors that influence the effectiveness of training in producing learning, performance, and transfer outcomes. Recent theory and research has resulted in models of training effectiveness that consider the large space of variables that mediate the influence of training on learning and transfer (e.g. Alvarez, Salas, and Garofano, 2004; Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991; Colquitt, LePine & Noe, 2000) and the categories of learning outcomes (e.g., Kraiger et al, 1993). These variables typically include individual factors such as experience level, abilities, trainee expectation and motivation; training factors, such as delivery mode, instructional methods, quantity of training, and feedback strategies; and organizational and transfer environment factors. In addition, the influence of these factors on the effectiveness of training can vary as a function of the tasks to be learned. Task characteristics include factors such as task type, task complexity, task performance requirements, and the types of knowledge and skills required. Of the factors that impact training effectiveness, the training community has more control over the training methods.

The overall goal of this research project is to understand how performance outcomes can be improved for different types of Army relevant tasks through the use of training methods tailored to the target task. In particular the research is aimed at identifying quantitative relations between training methods, task characteristics, and outcome measures that can provide an evidence base to guide training design decisions. For a given task, what training methods or combinations of methods are more effective in terms of performance, retention, and transfer outcomes and/or reduction in training time to achieve criterion outcome levels? A second goal is to improve predictive models in order to extend this understanding to new tasks.

A primary focus is on tasks with complex cognitive and interpersonal components that are relevant to Army tasks such as robotics control, intelligence gathering and analysis, command and control, and stability operations, as examples. Research on methods for training complex cognitive tasks emphasizes strategies for adapting the difficulty of training tasks and comparing the effectiveness of transfer to the target task. Strategies for modifying the difficulty of training (e.g., via "training wheels", error-prevention, part task, adapting) to manage cognitive load (e.g., Pass & van Gog, 2009) will therefore be a central research focus. Task factors such as task type and task complexity can mediate the impact of different training methods on training effectiveness. A second area of emphasis is research on training methods that manipulate training time and practice methods for different types of tasks.

Training factors to be considered include the primary mode or environment for delivering the training. These can include classroom, simulator, electronic media such as computer-based and web-based (distance learning), and field training. While meta analysis comparisons of classroom instruction with computer-based and web-based training indicate advantages for the electronic media (e.g., Wishler and Olson, 2003), it may be that the training methods and design are the critical factors and not the delivery media (e.g., Sitzman and Wishler, 2006). Reviews of the research will focus on comparing training delivery modes for different task types and comparing training strategies using different delivery modes.

Research Approach

Our research approach involves a five-year plan that starts with a systematic review and synthesis of the training research literature focused on investigating the relationships among training factors, task characteristics, and effectiveness measures. The emphasis is on individual as opposed to team or collective training. A range of task types will be considered with the primary emphasis on complex cognitive tasks. A parallel effort focuses on Army specific training needs, methods, and task types. Meta analysis methods are being used to compare quantitative research findings from studies that investigate selected training methods for different task types and complexity levels. This phase of the research identifies empirically defined links between key training methods, task factors, and performance measures. In addition, the research review is identifying training methods where additional experimental studies are needed to validate effects for different task types or evaluation criteria. These findings will be used to refine the research framework and define research questions, variables of interest, specific hypotheses and overall experimental design for the experimental phase of the research. Experimentation will focus on evaluating training effects identified through the literature review and meta-analysis in the context of specific Army relevant tasks. Given the emphasis on complex

cognitive tasks the experimentation phase is expected to focus on methods for enhancing training for complex tasks by modifying the difficulty of training tasks and evaluating transfer to a more complex target task. At the conclusion of the experimental phase, the existing research base and the sum of the experiments will provide data to better understand the relationships between training methods and performance outcomes. The data collected will be used to specify formal relationships between training methods and performance variables for different task types

Accomplishments

Accomplishments to date include a broad review of the training strategies literature, currently in progress, focused on the relative effectiveness of different training methods and the moderating effects of task and learner characteristics. A systematic selection, review and synthesis process has yielded over 400 review articles, meta-analyses, Army/military specific studies and research papers addressing the effectiveness of training methods and strategies. Papers were coded to identify training delivery methods, training strategies addressed, training task characteristics, individual factors, and outcome criteria. Findings are being sythesized by training method and task type to identify potential research gaps and areas for more comprehensive review and meta analysis.

In addition, a comprehensive meta analysis is in progress to investigate strategies for managing the difficulty of training task and identifies their relative effectiveness measured in terms of performance on a transfer task. The literature search and meta-analysis is focused on various methods for scaffolding, part task training, variable priority training, error prevention, increasing training task difficulty, adapting training to learner characteristics, providing learner control, guided exploration, simulation fidelity, scheduling practice and providing feedback. Approximately 60 studies meeting the training transfer requirements have been idetified for the first four training strategies.

A database repository was developed to organize, code, and synthesize the training effects knowledge derived from literature review and meta-analysis. The searchable database is populated as the literature review and the meta-analysis progresses and studies are coded. In addition to supporting near term coding and synthesis goals, the database structure provides a foundation for the longer term objective of prvoding a framework to identify appropriate training strategies for different combinations of task and individual factors based on available research results.

Contributions to Basic Research

Anticipated results from this project will contribute to basic research in a number of ways. Meta analysis results will provide a synthesis of the empirical studies comparing the effectiveness of selected training methods moderated by task factors and evaluation criteria. The research synthesis will also identify training methods and task areas where additional research is needed to evaluate effectiveness. The anticipated experimental results will extend the basic research to provide additional knowledge on the effectiveness of training methods for a range of task environments. The results of this effort will contribute to the empirical research predicting the impact of different instructional strategies and task characteristics on methods to accelerate

learning, maximize retention, and enhance training transfer. The application of the effect size data and the experimental data to predictive models provides a way of quantifying the effect of training factors on performance for different tasks.

Potential Army/Military Applications

The results of this research effort will provide a framework identifying effective training design methods for application to Army relevant tasks. A framework and model that provides heuristics and quantitative algorithms relating training methods to performance for different types of tasks could be transitioned into a number of products and widely applied in Army and other military training research and practice. For example, it has potential to provide training program managers and unit level training staff with guidelines for the best ways to develop or augment training for specific task types.

Future Plans

Research plans for this year include completing the systematic review of the training effectiveness research literature focused on review and meta-analysis studies and empirical studies investigating relationships among training factors, task characteristics, and effectiveness measures. This review will provide an overview of the current theoretical and operation frameworks for understanding the effect of training on performance. The literature review will include a summary of research studies categorized by: training mode, training strategies used, task type, evaluation criteria, and performance measures used. In parallel with the literature review, we plan to continue to investigate Army training methods and strategies currently in use to train different Army task and skill types and recent advanced training concepts. This task has involved interviews with Army training experts focused on training challenges, effective training approaches, and training strategies used from the perspective of current and evolving practice. The objective is to gain an understanding of Army training issues to ground the research by identifying the types of Army tasks that need to be studied and the training methods and strategies that are most meaningful to current training needs.

The initial meta-analysis of training methods that manipulate the difficulty of tasks during training will be completed this year. Preliminary findings have suggested a need for additional research on the effect of different training methods on transfer performance and on the moderating effects of task characteristics on performance. Over the next few months as the meta-analysis is completed, a research plan will be developed to conduct basic research studies that address identified research questions relevant to Army task domains and training needs.

Automatically-Learned Representation of Complex Task Environments to Support Training and Collaboration

Contract#: W91WAW-07-C-0013   **Contract Dates:** 04/09/2007 to 04/08/2010
Institution: Knowledge Analysis Technologies, LLC   **PI:** Peter W. Foltz

Problem(s)/Research Question(s) –
The problem motivating the project, which is related to the inherent complexity of distributed collaborative team environments, is that effective measurement goes beyond the powers of traditional task analysis. We are developing theory, methods and a generic automated tool that will model, assess, accelerate and aid the experience-based acquisition of human knowledge and skills needed in distributed collaborative tasks.

Technical Barrier(s) –
Conventional approaches to measurement in task environments require labor-intensive efforts in developing task, cognitive or user models. However, in complex team environments these models become burdensome to develop as well as quite fragile. This limits the ability to develop effective measures of team performance as well as systems to evaluate, train, or monitor teams.

Significance/Impact for Basic Research –
This project addresses two interlocking basic research problems. The first is how to use modeling approaches to improve the efficiency and effectiveness of distributed net-based collaborative problem solving and team performance operations. The second is to improve techniques for characterizing performance within complex problem solving situations.

Potential Transitions –
The potential for technology based on the underlying theories and methods developed under this project can lead to software for automated team and individual performance analysis within training systems. These can include leadership training systems such as those at the Army War College of Fort Leavenworth, automated After Action Review (AAR) systems for convoy training at Fort Lewis and the National Training Center, as well as for any command and control and/or complex team simulation training systems.

Overview

We are developing theory, methods, and a generic automated tool that will model, assess, accelerate and aid the experience-based acquisition of human knowledge and skills needed in distributed collaborative tasks. The automated task analysis tool relies on a new fundamental extension of the mathematical/computational theory *Latent Semantic Analysis* (LSA, originally invented by principals of the project) that is already in use in automated Army training, assessment and performance aids. However, instead of learning from massive bodies of expository text as in traditional LSA, the extension learns from extensive records of the sequential communications and actions of actual collaborating groups.

The technology is especially promising for tasks and situations where sufficient characterization of the context, knowledge and skills involved is beyond the reach of coding by

subject matter experts and where capture by manually constructed models is inordinately difficult because of the number and fine granularity of the variables involved as well as the rate at which they change. Some examples of such tasks are decision-making and planning by distributed command groups, team control of unmanned vehicles, aircraft control, and trouble-shooting of electronic weapon systems.

The machine-learning approach allows the system to emulate and predict both nearly repetitive and novel human actions in the same domain as that in which the system was trained. It does this by measuring the similarity of new events to old, where events are the combination of potentially infinitely variable situations with infinitely variable human knowledge, decisions and actions. It is able to use communication data (both content and patterns of who speaks to whom, when), participant actions (button presses, commands, decisions), and environment and system states.

To accomplish the task, we needed to obtain data from complex, problem solving or planning activities. The planned datasets were 1) asynchronous discussion data generated during an Interagency Process Simulation conducted at the Army War College, 2) communications and control actions made by teams of students at the Naval Postgraduate school using a command and control simulator called C3Fire, and 3) team communication and control actions collected at Fort Lewis using the Ambush! simulator for convoy operations training.

Success of the research endeavor is being evaluated by the ability of the system to dynamically predict the probability of outcomes of new trials as they progress, score individual actions according to their predicted effects thereon, assess the overall task-relevant knowledge and collaboration skills of both individual participants and groups as a whole, and to accelerate task success with automatically generated warnings, critiques and coaching of play. Such measures may be thought of as objective assessments of fundamentally subjective human abilities, and in most cases intuitive rather than logically reasoned because of the complexity of both the task conditions and interpersonal relations that are comprised of communication and influence factors.

The idea here is that such a modeling ability is applicable to instruction and aiding of real-world group tasks such as command, control and mission planning, where the complex background and context data for the machine learning model comes from human communication and system states in actual usage of simulator trainers or operational performance rather than (or in addition to) expertly constructed assessment and training systems.

Research Approach

Our goal is to research and develop methods for automated task and communication analysis and evaluate their predictive performance. The original research plan was as follows. Changes to the research plan are detailed in the accomplishment section:

Year 1
The initial focus of year one is on obtaining the datasets and initial development of the modeling technology. The key tasks to be accomplished in year one have been:
1) Obtain datasets. We will obtain data from complex events. We will further obtain subjective and/or objective performance measures associated with the data.

2) Develop technology for LPSA analysis of team interaction streams Then, using the available data, we plan to generate the high-dimensional LPSA problem spaces.
3) Further refine research methods for analysis of content of spoken and typed communication. Determine types of language indicative of good and poor performance (e.g., loss of situation awareness, planning, use of acknowledgements to indicate information received).

Year 2

1) Evaluate LPSA representations. We will iteratively refine the LPSA methods as well as test the boundaries of how much training data is necessary for effective and generalizable LPSA representations. Evaluation of performance of the LPSA models will be made against the objective and subjective team and individual performance ratings.
2) Integrate LSA-based communication predictions within LPSA model. We will develop and refine the methods for predicting important communication events and integrate them.

Year 3

1) Develop and evaluate an integrated LPSA-based moderation tool. The LPSA tool will incorporate data from the communication, action and event streams.
2) Tests of generalizability across datasets. Based on the analyses performed on the three diverse datasets, we will determine the most appropriate domains for the application of LPSA modeling.

Accomplishments

The research has been underway since April 2007. We have chosen to weave two interacting activities to achieve the goals of the project. The first is obtaining datasets amenable to this type of analysis. The second is investigating and developing representational techniques and applying these techniques to the datasets and finally evaluating their performance within the domain of the datasets. We have initially been investigating three classes of datasets: data from the C3Fire system, a firefighting simulation that is used in leadership training, multiplayer Internet games, and driving simulator data. Unfortunately, hoped for data from the convoy training simulation system Darwars Ambush! has not become available to us due to unexpected difficulties in pulling appropriate log data from the underlying simulation system.
C3Fire: The initial dataset we tackled consisted of 48 C3Fire missions that were collected as part of another project. The C3Fire data consisted of simulator log data from 8 three-person teams attempting to extinguish a forest fire in each of 6 scenarios. To apply LPSA to the C3Fire data required generalizing our software infrastructure to support LPSA, building on existing support for LSA. The major effort is in generating the equivalent of terms for event type data. As our understanding of the requirements of LPSA increase, we will continue generalizing the software to meet these needs.

We initially began exploring the C3Fire data, but it proved a tough domain to get results that were likely to tell us much about LPSA as a similarity measure. After further analysis, we chose to put this data aside and try our initial experiments with a simpler, larger dataset with better defined semantics and then to come back to the C3Fire data once we had a better established analysis procedures. The difficulties with this dataset included too few mission to perform effective dimension reduction, lack of an obvious scheme for developing terms, and an ill-defined semantics over groups of events. Learning from the complexity of this data the next domain we modeled was car simulator data, and our results with that dataset indicates that

13

revisiting the C3Fire data with our new found experience, especially if we can obtain more missions, will likely lead to useful models of that dataset.

Multiplayer Game Data: Before describing the car simulator data, we note that we have investigated the use of multiplayer Internet games as a source of task environments. Not having access to the Ambush! log data, led us to trying to obtain similar types of simulation game data. Both contacts within Pearson and with an online provider were initiated, but none of the first set of games examined had the face validity to fit our needs. Before we had traveled too far down this path, car simulator data from the Driving Wisdom project became available.

Driving Wisdom Car Simulator Data: One of the goals of the ARI-sponsored Driving Wisdom project is to investigate and elucidate the use of tacit knowledge in experienced drivers in order to provide Soldiers with training that improves their recognition of driving hazards. One experimental component of the project consisted of studying subjects performing a series of five driving scenarios conducted within a driving simulator. Both experienced and less experienced drivers were recruited for the study, and in addition a number of experimental conditions were manipulated among the drives. The variability of the drives and subjects provides meaningful subsets of the data on which to apply similarity measures. Simulator state data from the driving simulator was kindly made available to us by the principle investigators. In addition to the simulator data both objective and subjective performance data on aspects of the drivers' actions was collected. This dataset, by simplifying many aspects of the LPSA modeling process has turned out to be quite valuable in helping to develop the methodology of going from task environment data to analysis. We are currently working to analyze the simulator data, and evaluate the validity and usefulness of the types of analysis that LPSA can provide. The experimental design divided each drive into segments, where a segment consisted either of time between events or time during an event. An event consisted of a functional element of the drive such as receiving a cell phone call or having another car pull into an intersection. Each approximately 12-minute drive was divided up into approximately 15 segments.

Driving Wisdom Analysis: We had dual interlinking goals in the analysis of this data both proving the validity of the representation as well as viewing the simulator data from a unique perspective. Using the definition of a term as the composition of human controlled variables in a time-sliced simulator state, and segments as paragraphs, we built a semantic space. The space consists of 2597 terms and 1821 documents/paragraphs. For our initial run we chose a space with 100 dimensions. In the future we intend to explore other dimension sizes to see the impact on the performance of the space. Given this semantic space, the next step is to use it to build vectors of meaningful chunks and compute their similarity.

The first obvious semantic unit is to compute a vector representation for each drive, and then compare the similarities over various groupings of drives. The drive is one of the units of analysis for the Driving Wisdom project and making similarity statements about subsets of the drives is a relevant unit of analysis. The first comparison considered is whether the experienced driver group is more coherent (had on average higher cosine among the drives) than the less experienced group. Using the new semantic space the cosine between experienced group drives (M = .42, SD = .21) is significantly higher than the cosine between drives from the less experienced group (M = .37, SD = .24), with t(3400) = 6.0, p = 2.8e-9. Also note that the dispersion of the experienced group with standard deviation of .21 is smaller than that for the less experienced group with .24. This result provides additional support for the notion that there

are fewer ways to correctly execute the drives as shown by the experienced group than there are less correct drives as shown by the less experienced group.

In a second analysis, an informal observation was that the lead vehicle condition is substantially more constraining on the driving behavior than the non-lead vehicle condition. We excluded drive 5 in this analysis, since it had additional events that would likely confuse the analysis. The constraining nature of the lead vehicle intuition is supported by the data and this similarity metric, where the cosine between drives in the lead vehicle condition (M = .64, SD = .15) was significantly greater than the non-lead vehicle condition (M = .43, SD = .26), with t(1800) = 23.6, and p < 2.2e-16.

One of the findings using the performance measures is that of the two distraction conditions, either a simulated cell phone call or a passenger condition, the passenger condition was found to more strongly negatively impact a search task the subjects were asked to conduct. In contrast to these findings, there was not a significant difference in the coherence of the groups between the passenger condition (M = .38, SD = .22) and the cell condition (M = .39, SD = .24) , t(3400) = 1.33, and p = .18. It may require a more subtle analysis of between group coherence differences to detect this type of performance impact, and exploring this issue is in our future plans.

For the last analysis, we wanted to confirm the observation that drive 5 with its unexpected events and less constrained instructions would be less coherent than the other drives. The pairwise cosine similarity for the other drives (M = .42, SD = .24) is significantly more coherent than for drive 5 (M = .36, SD = .16), t(315) = 5.8 and p = 1.6e-8 confirming this observation.

There are a number of other groupings of the data that we hope to explore. We have discussed these results with one of the principle investigators of the Driving Wisdom project, and she feels these results support their hypothesis that tacit knowledge is important in driving expertise and that this type of analysis may be helpful in their continuing investigations.

The above analysis is at a quite coarse level, looking at the coherence of a large number of drives. Almost at the other end of granularity scale is a k-near analysis. Since we are building a similarity measure, a k-near analysis looks at the k nearest drives to a given drive, and tries to understand the qualities of that set. We have just started this analysis and early results indicate some differences from communication data in that drives seem to influence k-near more than driver, where the opposite is more common with communications. We are working to obtain additional k-near insights. The final type of analysis is to see how well the similarity measure allows predicting similar performance. The idea here is that the performance of similar drives should be similar. It isn't entirely clear that will be true for the objective measure of this dataset since it isn't clear how tied driving and the objective search task will turn out to be, but we will try predicting the objective measure and the subjective measure. We are just starting this work and it will require some additional software changes, as we need to generalize our more language oriented features toward this event and state type data.

A second area that we have been investigating is using technologies beyond LPSA to provide similar levels of analytic power as LSA. For our work with Darcat convoy data we had begun to explore Stochastic Process techniques such as moving windows. We are planning on generalizing this work to use Hidden Markov Models (HMMs) to model the simulator data. The difficulty in building HMMs to model driver behavior is that during a drive there are often

context switches (such as the events in the Driving Wisdom driving scenarios) that have been difficult to model with plain HMMs. The driver modeling community has encountered these issues and a number of promising generalizations of HMMs have been emerging, such as Stochastic Switched models that can automatically switch HMMs depending on various external conditions.

Contributions to Basic Science

The project is directed at two interlocking problems. The first is to improve the efficiency and effectiveness of distributed net-based collaborative problem solving and team performance operations. The second is to improve techniques for characterizing performance within complex problem solving situations. These problems have traditionally been approached by task analysis methods based on interviews with Subject Matter Experts (SMEs), observations of expert performance and manually constructed specifications and organization of tasks and their associated required knowledge bases. The innovative approach responds to the burdensome time, expense and intrinsic limitations of conventional techniques when the environmental context is so complex—composed of such a too-large or an infinite combination of influences—as to profit from the application of automatic computational techniques. There are two critical differences in the approach relative to other approaches to modeling complex tasks. They are 1) We capitalize on large amounts of team data, including analysis of the content of what is being expressed, the patterns of communications and actions of participants. 2) The techniques for characterizing performance permit automated derivations of task analyses without tedious hand-coding of tasks.

Potential Army/Military Applications

Monitoring, assessing and moderating teams of decision-makers in complex military environments require effective tracking of individual and team performance. The research advances our understanding of modeling performance in complex task environments and permits analyses of complex communication and action patterns. It can therefore be applied in any area where automated task analysis and performance measurement is needed in interactive team situations.

The potential for technology based on the underlying theories and methods developed under this project can lead to software for automated team and individual performance analysis within training systems. The toolset can show integrated concepts and approaches to analyzing communication and control actions, providing predictions of performance and automated feedback. Such a system could provide automated approaches to:

- Detecting individual and team errors and knowledge gaps
- Predicting cognitive workload and situation awareness
- Predicting overall team performance
- Predicting failures in team process and inappropriate contexts for the mission
- Monitoring group dynamics
- Providing automated AARs
- Assessing and training leadership

Future plans

In the next period, we plan to continue analysis of the Driving Wisdom data, especially k-near relationships and attempt to predict various performance metrics using the new similarity

measure. This will require some additional research to better understand appropriate features for modeling, and also software modifications to support those features. In the spring, we are also expecting additional simulator data from new Driving Wisdom experiments that are in the planning states. Depending on availability, we may go back and attempt to reanalyze the C3Fire data we have or attempt to obtain additional C3Fire data to analyze. We also plan to have implemented some of the Stochastic Process Analysis software and model some of existing data. Finally, we will work to identify DOD entities that could benefit from the automated modeling and performance prediction technology. We would work with them to develop a more robust toolset and integrate the system in either their live or virtual environment training.

Improving Coping by Building Self-Control

Contract #: W91WAW-08-C-0032　　　　　**Contract Dates:** 09/01/2008 to 08/31/2011
Institution: University of Albany　　　　　**PI:** Mark Muraven

Problem(s)/Research Question(s) – How does self-control contribute to coping outcomes? Is it possible to improve self-control to lead to better coping with stress?

Technical Barrier(s) – Research on the application of self-control to coping is not well developed. Interventions to build self-control are primitive or non-existent.

Significance/Impact for Basic Research – A better understanding of the role of self-control in coping with stress would help advance theories of both coping and self-control. Discovering how to improve self-control may also lead to increased knowledge of the operation of self-control.

Potential Transitions – Knowledge gained in this basic research may be useful in:
- ARI's applied training programs
- ARI's Leadership Development program, as well in development of selection, classification, and performance metrics
- Army Behavioral Health Initiatives

Overview

Coping with stress should require self-control. Successful coping requires the individual to regulate him/her self in order to relieve emotional distress, overcome obstacles, negate negative thinking, and stay on track toward achieving his or her goals. This suggests that coping requires the individual to stop an emotion, override his or her thoughts, and restrain impulses to give up. Indeed, Eisenberg et al. (1997) suggest "the ability to inhibit behavior appears to be especially linked to quality of coping" (p. 55).

However, to date, no research has examined this important issue. This is crucial, because knowing the role of self-control in coping with stress may help in the understanding of how coping operates. Moreover, at an applied level, this knowledge may lead to programs that improve coping with stress and predict when coping is likely to breakdown.

In addition, our understanding of how self-control itself operates is still rudimentary. Muraven and colleagues (Muraven & Baumeister, 2000; Muraven, Collins, Shiffman, & Paty, 2005; Muraven & Slessareva, 2003; for a review, see Muraven & Baumeister, 2000), have developed a model of *self-control strength.* According to this model, any and all attempts at self-control draw upon and require the use of self-control strength. Thus, coping with stress, as an act of self-control, should require self-control strength. Individuals higher in strength should be better able to exert self-control than individuals lower in strength.

This pool of strength may change in response to both short-term and long-term self-control efforts, however. The act of exerting self-control is theorized to deplete some of this strength. Because self-control strength is limited and replenished slowly, individuals who exerted

self-control should be lower in strength than individuals who did not exert self-control. This leads to Specific Aim 1:

- *Specific Aim 1: Demonstrate that participants' level of self-control strength is related to their ability to cope with stress in both the laboratory and real world.*

There is a second aspect of the strength model. The model predicts that, over time and with rest, the exertion of self-control should increase individuals' capacity for self-control. Much like physical exertion leads to weakening and fatigue (depletion of strength) in the short-term but an increase in strength and greater resistance to fatigue in the long-term (expansion of strength), the exertion of self-control may have distinct long-and short-term effects. Immediately after exerting self-control, strength is depleted and self-control performance suffers. With rest and enough time to recover, however, strength is regained (Tice, Baumeister, Shmueli, & Muraven, 2007) and possibly even enhanced (Muraven, Baumeister, & Tice, 1999).

Hence, it should be possible to improve individuals' ability to exert self-control by having them practice small acts of self-control for several weeks. This increased self-control may be reflected in better coping outcomes (see Figure 1). This is Specific Aim 2:
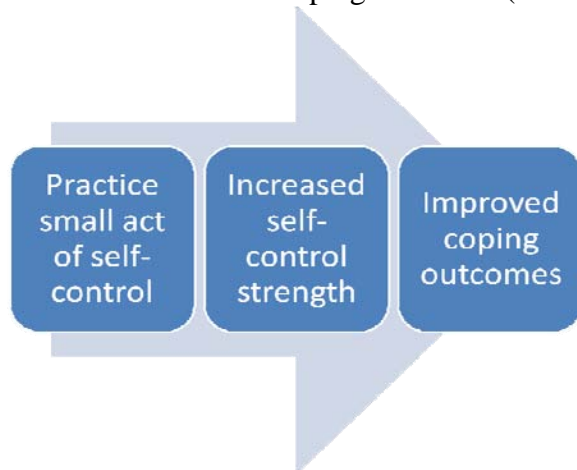
Figure 1

- *Specific Aim 2: Demonstrate that practicing small acts of self-control lead to increased self-control performance, which results in better coping outcomes.*

Research Approach

The present research examines how self-control affects coping with stress. This relationship is being assessed using both laboratory and field (EMA or ecological momentary assessment) data collection. In particular, a sample of 350 college age participants' responses to a laboratory stressor is being measured two times, each session separated by two weeks, using a variety of self-report, behavioral, and psychophysiological measures. Participants are randomly assigned to exert self-control (or to neutral condition that is matched in difficulty and frustration to the self-control task but does not require self-control), to assess the role of self-control strength in coping outcomes.

Between the laboratory sessions, participants are carrying a PDA (personal digital assistant) for four weeks as they go about their daily life. They are using this to answer questions about stressors, their coping with these stressors, and self-control demands and level of strength in the real world. In this way, the relationship between self-control strength and coping outcomes can be measured in day-to-day living.

In addition, between the first and second laboratory sessions, participants are instructed to practice a task that requires self-control, such as avoiding sweets or cutting back on cursing (participants in the control condition are practicing tasks that do not require self-control but that are matched in difficulty, frustration, effort, and self-awareness to control for self-monitoring and self-efficacy). This task should increase participants' self-control and lead to better coping. Participants' coping after practicing will be compared to their coping before practicing, as well as to participants who practiced tasks that should not increase self-control strength. I hypothesize that participants who practiced self-control should exhibit better outcomes in both the laboratory and diary measures of coping.

Accomplishments

Data collection is ongoing and preliminary results look promising. It appears that self-control is related to psychophysiological reactions in the laboratory. Individuals lower in self-control exhibit greater heart rate reactivity as compared to individuals whose self-control capacity is not diminished. We are still collecting data on stress reactions in the field, as well as whether it is possible to build self-control strength to improve coping outcomes.

Contributions to Basic Science

This project may help in understanding the role of self-control in coping with stress. To date, the literature of coping is disjointed, with an unclear picture of how coping works, why it sometimes fails, and how it can be improved. The results from this experiment may help develop a more comprehensive model of coping has self-control at its core. Such a model, when tied into the developing research on self-control, may suggest ways to improve coping and prevent stress related problems.

In addition, this research will help advance theorizing on self-control. In particular, this research will examine the role of self-control strength in self-control performance. It could lead to better measures of self-control, as well as a clearer picture of how self-control demands affect subsequent self-control performance in the real world.

Finally, this project will increase our knowledge about the development of self-control. Preliminary work has suggested that it is possible to improve self-control capacity by practicing small acts of self-control regularly. The present research should refine these ideas to examine boundary conditions as well as alternative models, to better determine how and why practicing self-control works.

Potential Army/Military Applications

The theory, if supported, suggests a critical role for self-control in coping with stress. Insofar as modern military missions are stressful and require coping, individuals who are better able to cope may function at a higher level. Hence, Soldiers with greater self-control capacity may make better decisions under stressful circumstances, be more flexible in changing environment, and remain focused on the mission goals in difficult situations. The military may find it useful and cost effective to assess Soldiers for self-control capacity, to use in personnel decisions, as well as in behavioral health interventions.

Moreover, the results may suggest economical ways to improve self-control. Such an intervention, either Army-wide, or targeted to specific individuals, could greatly enhance the effectiveness of the military.

Future Plans

Data collecting is ongoing. The preliminary data from the laboratory measures psychophysiological reactions to stress looks promising. We hope to submit some early findings to conferences this year.

References

Eisenberg, N., Fabes, R. A., & Guthrie, I. K. (1997). Coping with stress: The roles of regulation and development. In S. A. Wolchik & I. N. Sandler (Eds.), *Handbook of children's coping: Linking theory and intervention* (pp. 41-70). New York: Plenum.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*, 247-259.

Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. *Journal of Social Psychology, 139*, 446-457.

Muraven, M., Collins, R. L., Shiffman, S., & Paty, J. A. (2005). Daily fluctuations in self-control demands and alcohol intake. *Psychology of Addictive Behaviors, 19*, 140-147.

Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29*, 894-906.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology, 74*, 774-789.

Tice, D. M., Baumeister, R. F., Shmueli, D., & Muraven, M. (2007). Restoring the self: Positive affect helps improve self-regulation following ego depletion. *Journal of Experimental Social Psychology, 43*, 379-384.

Combating Attrition from Online Instruction

Contract #: W91WAW-09-C-0086   **Contract Dates:** 8/04/2009 to 8/3/2011
Institution: Advanced Distributed Learning   **PI:** Traci Sitzmann
                                                  Co-PI: Katherine Ely

Problem(s)/Research Question(s) – Why do people drop out of voluntary self-development courses? How can we reduce attrition from online training?

Technical Barrier(s) – Attrition from organizational and military training courses has largely been ignored in the research literature. There are no existing models that explain why trainees drop out of online work-related training and there is a dearth of research on interventions that can be implemented to reduce attrition.

Significance/Impact for Basic Research – This research will permit us to begin theory development surrounding the training attrition process and to develop theoretically based interventions that can be implemented in online training to reduce attrition.

Potential Transitions – This basic research may be employed in a breadth of Army online courses in order to enable trainees to complete training and master the course content. Specifically, the interventions that are investigated could be implemented in courses that traditionally have high attrition rates and poor learning outcomes (e.g., Army Pharmacy Technician—68Q—training).

Overview

Numerous observers have noted that attrition may be problematic in online courses (e.g., Rossett & Schafer, 2003; Welsh, Wanberg, Brown, & Simmering, 2003), but our understanding of the factors that influence attrition in online work-related training remains limited. Although there are several existing models of the student attrition process (e.g., Bean, 1980; Spady, 1970; Tinto, 1975), most were developed to explain attrition from traditional classroom instruction. Furthermore, they focused on attrition in college settings and include commitment to the academic institution and social integration factors, which may not be relevant in asynchronous, organizational training courses.

More recently, researchers have expanded these models to better understand attrition from online courses (Berge & Huang, 2004; Boyles, 2000). Berge and Huang (2004) identified three categories of variables that influence attrition from online training: circumstantial (e.g., instructional design), personal (e.g., trainees' individual differences), and institutional (e.g., organizational values). Survey results generally support this model, with the majority of reasons provided for dropping out falling into the circumstantial or personal categories (Frankola, 2001; Muilenburg & Berge, 2005; Wang, Foucar-Szocki, Griffen, O'Connor, & Sceiford, 2003).

However, only a couple of studies have empirically examined predictors of attrition from military training. In 1979, Mobley and colleagues examined predictors of attrition from an 11 week Marine Corp recruit training program. They found that trainees were more likely to

complete training when: a) they entered training with both high expectancy and high intentions for completing the program; b) they thought their chance of finding an acceptable civilian role was low; c) they had more education before entering the program; and d) believed that their Marine role would lead to positive rather than negative outcomes. More recently, Fulton, Starnes, Caouette, Whittaker, and Ivanitskaya (2008) examined the factors that predict attrition from the Army Pharmacy Technician course. The mean graduation rate among 463 trainees from 2005-2007 was 71.3%. The probability of completing training was greater when Soldiers had higher ST ASVAB scores and a higher pay grade. Furthermore, both those who enrolled normally and due to recycle were more likely to complete training than those who enrolled because of reclassification of their military occupational specialty.

The current study will expand on this research stream to examine both trainee characteristics and the factors that occur during training (e.g., effort, course satisfaction, the extent to which their expectations were met, performance in the course) that lead up to trainees giving up and dropping out of the course. It is possible that trainees review the material, encounter a concept they do not understand, and drop out when they discover that increasing their concentration is not enabling them to understand the course material. However, boredom may also lead to attrition and trainees may drop out at the point when their concentration is waning.

Once we understand the factors that predict attrition, the next advancement in research should involve reducing attrition. Some organizations have attempted to tackle this problem by providing financial incentives for completing training and tying course completions to job promotions (Welsh et al., 2003). However, prompting self-regulation is much less complicated to execute and can be implemented free of charge. Prompting self-regulation involves asking trainees questions regarding whether they are setting goals, monitoring their learning behaviors, and implementing strategies to reach their goals. Sitzmann and Ely (in press) investigated the effect of prompting self-regulation on learning and attrition with 479 adults participating in a four-hour voluntary online course that was divided into four training modules. For example, trainees were asked, "Am I focusing my mental effort on the training material?" and "Are the study strategies I am using helping me learn the training material?" Attrition was 17 percentage points lower among trainees who were prompted to self-regulate than trainees in the control condition. Thus, this intervention substantially increased the probability of completing a voluntary online course. Prompting self-regulation also resulted in a 5 percentage point increase in test scores, relative to the control. On average, trainees in the control condition scored a 76% on the four exams in the course—which is a C average. Trainees who were prompted to self-regulate scored in the B range on the exams with an average test score of 81%. This result is consistent with several previous research studies that demonstrated that learning improved over time when trainees were prompted to self-regulate, relative to the control condition (Berthold, Nückles, & Renkl, 2007; Hübner, Nückles, & Renkl, 2006; Sitzmann, Bell, Kraiger, & Kanar, 2009). Furthermore, Sitzmann and Ely found that trainees who were prompted to self-regulate spent additional time reviewing the training material rather than merely clicking through the slides to reach the end of the course. Trainees who were prompted to self-regulate spent an average of 21 more minutes reviewing the course material per module than trainees in the control condition, which explains why they also learned more in the course.

This research will expand upon previous prompts research by examining the process by which prompting self-regulation influences attrition from online training. One of the factors that leads to high attrition in online training is lack of an internal locus of control (Martinez, 2003). Trainees have an internal locus of control when they believe that their performance is under their control rather than being controlled by external forces and events (Rotter, 1954). Prompting self-regulation may induce a state internal locus of control by informing trainees that they have control over their performance in the course. Thus, we will empirically investigate if prompting self-regulation induces a state internal locus of control and whether this mediates the effect of the intervention on attrition from voluntary online training.

Furthermore, we will develop and test a second intervention that can be implemented in online training to reduce attrition. The second intervention involves having trainees develop a plan at the beginning of the course for when and where they are going to complete training.
Phase models of self-regulation suggest that creating a plan is a critical first step for engaging in self-regulation (Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman, 2000). We believe that one of the primary reasons that trainees do not complete online courses is because they become busy and fail to set aside the time required to complete the course. Creating a plan should help ensure that trainees set aside sufficient time for participating in the course and contemplate how they are going to eliminate distractions that may prevent them from concentrating on training. Via creating a plan and following through on the plan that is established, we are hoping to reduce procrastination and, ultimately, enhance learning and reduce attrition from online training.

Research Approach

Advertisements will be posted on Google and Yahoo for free online Microsoft Excel training in exchange for research participation. Trainees who sign up will participate in an online training course and will complete a pretraining measure to assess their individual differences (e.g., demographics and trait persistence). Moreover, they will complete study measures several times during training in order to assess their self-regulatory processes and will complete an exam at the end of each of the modules. The learning management system will capture the amount of time spent in each of the training modules, the dates for when trainee login to the course, and attrition from training.

Accomplishments

Two months after our kickoff meeting with ARI, we have already put our experiments online, created a website for data collection, implemented our recruitment plan, and enrolled 635 people in the experiments. Moreover, we are developing a learning management system to capture behavioral indicators of self-regulation in future research endeavors.

Contributions to Basic Research

This research will provide preliminary information that can be used to develop a theoretically driven model of the factors that predict attrition from voluntary online training. It will also investigate interventions that can be implemented in online training, their effects on learning and

attrition, and the self-regulatory processes that mediate the effects of the interventions on these training outcomes.

Potential Army/Military Applications

The Army is utilizing online training to increase personnel readiness. Online training enables the military to provide just-in-time training anytime and anywhere. However, attrition from training greatly decreases military benefits and increases the cost of training per person who completes the course. For example, Fulton and colleagues (2008) estimated that the cost per student who drops out of the Army Pharmacy Technician training is $50,000 (including salary, transportation, and material costs). Thus, the annual cost of attrition from this course is $3.73 million (given 28.7% of 260 students per year drop out). Across all Army training programs, the cost of attrition from training is staggering.

In order for the Army to reduce attrition, they must first identify who is likely to drop out and the factors that lead to attrition from online instruction. This research will lead to the development of a model of trainee characteristics and processes that occur during training that lead to high attrition rates. Furthermore, two no cost interventions will be developed and empirically tested in terms of their effectiveness for reducing attrition and enhancing learning. Reducing attrition from online instruction will ensure that the Army is mission ready and Soldiers have the skill sets necessary to successfully perform their job duties.

Future Plans

The attrition domain is wide open given the dearth of research in the area. Thus, future projects will investigate additional factors that predict attrition, including individual differences and self-regulatory processes during training. We will also investigate how these factors can be mitigated in online training. For example, control theory proposes that if a person encounters an obstacle during goal pursuit, the interruption triggers a reassessment of the situation (Carver & Scheier, 1990, 2000). The reassessment leads trainees to examine the probability that they can reach their goals if they invest additional effort. If trainees are unaware of strategies that they can use to succeed in training, they physically and mentally disengage from training. However, these theoretical assumptions have not been subjected to empirical scrutiny. Thus, one possible research project would be examining the obstacles that trainees encounter during training and their reasoning process in deciding whether to remain in training or drop out. Once the process of attrition is understood, we could begin to implement interventions at the points in training when trainees need cognitive and emotional support to succeed.

References

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12,* 155-187.

Berge, Z., & Huang, Y. (2004). A model for sustainable student retention: A holistic perspective on the student dropout problem with special attention to e-learning. *DEOSNEWS, 13*.

Retrieved March 02, 2009 from
http://www.ed.psu.edu/acsde/deos/deosnews/deosnews13_5.pdf.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17,* 564-577.

Boyles, L.W. (2000). *Exploration of a retention model for community college student.* Unpublished doctoral dissertation, The University of North Carolina at Greensboro.

Carver, C. S., & Scheier, M. F. (1990). Origins and functions of positive and negative affect: A control-process view. *Psychological Review, 97,* 19-35.

Carver, C. S., & Scheier, M. F. (2000). On the structure of behavioral self-regulation. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation.* New York: Academic Press.

Frankola, K. (2001). Why online learners drop out. *Workforce, 80,* 53-58.
Fulton, L., Starnes, W., Caouette, M., Whittaker, D., & Ivanitskaya, L. (2008). Explaining and forecasting attrition in the Army Pharmacy Technician course. *Military Medicine, 173,* 1219-1224.

Hübner, S., Nückles, M., & Renkl, A. (2006). Prompting cognitive and metacognitive processing in writing-to-learn enhances learning outcomes. In R. Sun, N. Miyake, & C. Schunn (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 357 - 362). Mahwah: Erlbaum.

Martinez, M. (2003). High attrition rates in e-learning: Challenges, predictors, and solutions. *The eLearning Developers Journal.* Retrieved March 09, 2009, from http://www.elearningguild.com/pdf/2/071403MGT-L.pdf

Mobley, W. H., Hand, H. H., Baker, R. L., & Meglino, B. M. (1979). Conceptual and empirical analysis of military recruit training attrition. *Journal of Applied Psychology, 64,* 10-18.

Muilenburg, L.Y., & Berge, Z. L. (2005). Student barriers to online learning: A factor analytic study. *Distance Education: An International Journal, 26,* 29-48.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). New York: Academic Press.

Rossett, A., & Schafer, L. (2003). What to do about e-dropouts. *Training and Development, 57,* 40-47.

Rotter, J. B. (1954). *Social learning and clinical psychology.* Englewood Cliffs, NJ: Prentice Hall.

Sitzmann, T. M., Bell, B. S., Kraiger, K., & Kanar, A. M. (in press). A multilevel analysis of the effect of prompting self-regulation in technology-delivered instruction. *Personnel Psychology.*

Sitzmann, T., & Ely, K. (in press). Sometimes you need a reminder: The effects of prompting self-regulation on regulatory processes, learning, and attrition. *Journal of Applied Psychology*.

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1,* 64-85.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45,* 89-125.

Wang, G., Foucar-Szocki, D., Griffen, O., O'Connor, C., & Sceiford, E. (2003). *Departure, abandonment, and dropout of e-learning: Dilemma and solutions* (Final research report to the E-learning Consortium, the Masie Center). Harrisonburg, VA: James Madison University.

Welsh, E. T., Wanberg, C. R., Brown, K. G., & Simmering, M. J. (2003). E-learning: Emerging uses, empirical results and future directions. *International Journal of Training and Development, 7,* 245-258.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2000). Attainment of self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13-39). San Diego, CA: Academic Press.

BRU RESEARCH OBJECTIVE #2: PROVIDE FUNDAMENTAL KNOWLEDGE TO IMPROVE LEADER AND TEAM PERFORMANCE

Research under this objective provides concepts and methods for accelerating leader development, understanding and developing leader adaptability and flexibility, and discovering and testing the basic cognitive principles that underlie effective leader-team performance. Understanding the dynamics of small group leadership in face-to-face and distributed team environments is critical to this research objective as well.

A 10-Year Extension to the Baseline Officer Longitudinal Development Study

Contract #: W91WAW-09-C-0087   **Contract Dates:** 7/20/2009 to 7/19/2012
Institution: University of Washington   **PI:** Bruce J. Avolio

Problem(s)/Research Question(s) – What are early predictors of the most ethical, authentic, adaptive and effective leaders trained by the U.S. Military Academy, and how do these graduates grow as leaders over time in terms of their leadership and performance?

Technical Barrier(s) – Obtaining a large sample of participants who are on a similar leadership development trajectory and that have similarity of experience has been a challenge that ARI and the USMA overcame with the original study. As such, this kind of extensive data collected over a long period of time on leadership development has not previously been collected, analyzed and reported on. Most leadership research has not been able to examine development over time with a unique cohort of military leaders.

Significance/Impact for Basic Research – Identifying individual characteristics, early life events/experiences, training components of the USMA training experience, and life/career experiences that occur post-graduation which most effectively impact leadership effectiveness and style in both civilian and military organizations over time.

Potential Transitions

The Army will obtain detailed information about what predicts the development of effective officer leaders over time. This will include:

- How intellectual ability, personality, self-confidence, logical reasoning and other personal characteristics predict changes in self-awareness, self-identity and leadership style in USMA and after graduation in the active military service.
- How leaders interact with superiors, colleagues, and followers, and the impact those interactions have had on their leadership development, potential and performance.
- How each of the above predicts leadership effectiveness for those BOLDS participants who have left the military and are in the private or public sector.

Overview

This project addresses several key issues regarding the development of adaptive leadership. First, we examine early antecedents that predict the later growth, development and performance effectiveness of leaders. These include leadership roles, experiences, and positions, as well as a range of individual differences such as personality traits, ability and academic performance. Second, we will examine how individuals grow and develop over time as leaders by focusing on events that may have contributed to such development and growth early to later on in the career. Finally, we will examine how change occurs in people over time on important dimensions such as leadership style, leadership motivation and leadership performance.

Over the last few decades there has been a growing interest in examining styles of leadership that are more in line with the dynamic contexts in which leaders operate. These more

dynamic orientations fall under the heading of *authentic transformational leadership* (Avolio, 2007). The interest in this kind of leadership ranges from understanding what predicts adaptive forms of leadership to understanding how it can be developed. In line with these interests, the BOLDS project will extend a longitudinal investigation begun in 1994 at USMA with a full class of cadets starting training (n=1143). The project is focused on predicting leadership style, change, growth and performance over a 10-year plus time span post-graduation. Our strategy involves analyzing existing data collected from the class of 1998, and collecting new data to track leadership development and growth over a full 15-year time span (e.g., including the 4 year USMA experience).

Previous data collected for BOLDS includes individual skills and characteristics, academic record, early leadership/life experiences, and leadership style exhibited by the cadet while enrolled at USMA. Data collected after their graduation includes military experience, positions held, and leadership style (which might change over time). Because of the longitudinal nature of this study, opportunities exist to assess growth, change and development in leadership over time.

One of the unique capabilities of this project will be the attempt to measure *extreme* performance, in addition to *typical* or *average* performance. There are relatively few leadership studies that examine what predicts extreme performance and the ways to best measure this. With the BOLDS data set, performance over extended periods of time are available, which can potentially allow us to differentiate leaders at the highest ends of performance domains, e.g., performing in extremely challenging contexts. The BOLDS project is in the position to make a meaningful contribution to the understanding of what facilitates extreme performance.

Another unique feature of the BOLDS project is that all participants started their careers at the same time in the same institution and all entered into the same ranked leadership position. There is no other longitudinal study in the field of leadership with such a controlled starting point for assessing leadership emergence, development and performance. Because of this structure, the BOLDS project can examine a wide range of contributors to leadership development that are not normally possible due to the variation in timing, training, and entry level positions that typically exist in a study sample. This is the reason that one of the goals of the project is to enlist the BOLDS participants to continue their participation in this study so that we may further examine their leadership growth and potential ten years after graduation from USMA.

Research Approach

The first task of the project is to compile all the data that has previously been collected on the BOLDS study participants. This includes information collected as part of their admission to USMA, their experiences and performances during their four years at USMA, and data collected by several different researchers after their graduation. The data needs to be systematically aggregated into a single repository so that a complete examination can occur of the information collected thus far on the BOLDS participants. It is expected that this aggregation will allow the researchers to analyze the data in a more rigorous manner than has previously been possible. While the primary goal of this part of the project is to confirm previously understood predictors

of leadership development in this group, it is anticipated that deeper and more complex relationships in the data may be revealed.

Upon confirming some critical relationships in the pre-existing data, we expect to be able to provide BOLDS participants with some idea of what variables predicted effective leadership in their cohort during their time in the Academy. We then hope to recruit the BOLDs participants into continued involvement with the longitudinal project and a new wave of data collection by providing over time further insights into their development as leaders

Accomplishments

As of this writing, the project is still going through the Institutional Review Board approval process to allow the investigators to begin. This review process will ensure both ARI and the University of Washington comply with the federal regulations associated with research involving human subjects.

Contributions to Basic Research

We expect to contribute to the understanding of what helps develop effective, ethical, adaptive and authentic leaders in the military (and potentially outside for those BOLDS participants who have left the military), and to understand what kinds of experiences have contributed to this development over time. This study provides us with a rare opportunity to actually look at the antecedents to leadership growth and performance.

Potential Army/Military Applications

We expect that the BOLDS project extension will help the U.S. Army identify unique development opportunities that could help predict ethical, effective, adaptive and transformational leaders. We also expect that the new BOLDS data collection will help identify critical life events that have occurred either within the military experience, or outside of the military, that have directly influenced leadership growth and development. Lastly, it is hoped that the new BOLDS data collection will help identify personal characteristics and/or types of events that contribute to successfully performing in extreme performance contexts.

Future Plans

Upon final Institutional Review Board approval to begin the project, the next year will be focused on compiling all previously collected data on BOLDS participants and generating a single data file that will permit analysis of the leadership development model that will be further examined in the new BOLDS data collection phase.

Leadership Development of Women and Ethnic Minority ROTC Cadets: The Potential Barrier of Stereotype Threat and Resiliency as a Buffer

Contract #: W91WAW-08-C-0037
Institution: California State Polytechnic University, Pomona

Contract Dates: 09/01/2008 to 02/28/2010
PI: Bettina Casad

Problem(s)/Research Question(s) – Are women and ethnic minority ROTC cadets susceptible to poor leadership performance due to stereotypes about their group's leadership inferiority? How does this stereotype threat affect cadets' leadership performance, emotions, and physiological reactivity? What individual difference variables predict resiliency to lower performance and unhealthy physiological responses?

Technical Barrier(s) – No research on stereotype threat has been conducted with a military population. Little research has examined the effects of stereotype threat on women's and ethnic minorities' leadership performance.

Significance/Impact for Basic Research – The research will advance theory on stereotype threat and link stereotype threat to physiological reactivity.

Potential Transitions – Knowledge gained in this basic research may be used in:
- ARI's applied training research programs (e.g., Future Force Warrior Training, Managing Training and Providing Feedback within Networked Embedded Training)
- ARI's applied selection research programs (e.g., SELECT21).

Overview

Stereotype threat is an emotional, cognitive, and physiological state that people may experience when they are concerned that their behavior or performance will be judged according to a negative social stereotype about a group to which they belong (Steele & Aronson, 1995). When under stereotype threat, individuals who highly identify with the particular domain are more likely to show performance decrements, even if they are highly skilled in the domain (Steele, 1997). For example, a widely held stereotype related to leadership is that women and ethnic minorities are not "leadership material," a stereotype that has been documented in military (Biernat, Crandall, Young, Kobrynowicz, & Halpin, 1998; Boldry, Wood, & Kashy, 2001; Boyce & Herd, 2003; Eagly, Karau, & Makhijani, 1995; Mettrick & Cowan, 1996) and non-military settings (Eagly & Karau, 2002; Davies, Spencer, Steven & Steele, 2005). Knowledge of this common cultural stereotype is enough to bias people's perceptions and behavior toward the stereotyped group (Devine, 1989), that is women and ethnic minority military leaders. Thus, women and ethnic minorities being evaluating on their leadership ability may have physiological reactions that reveal a state of threat (Blascovich, Mendes, Hunter, & Lickel, 2000) and disruptive thoughts (Cadinu, Maass, Rosabianca, & Kiesner, 2005) due to the heightened anxiety that they may be judged by the stereotype that women and ethnic minorities do not make good leaders.

Little research has been conducted on the effects of stereotype threat in a military context. Most of the military research has been conducted with military personnel regarding stereotypes about women or ethnic minority leaders (Biernat, Crandall, Young, Kobrynowicz, & Halpin, 1998; Boldry, Wood, & Kashy, 2001; Boyce & Herd, 2003; Eagly, Karau, & Makhijani, 1995), but has not examined effects of stereotype threat for women and ethnic minority military personnel themselves. What is known is that women and ethnic minority leaders are consistently devalued. For example, Biernat and colleagues (1998) had Army officers attending a leadership training course complete leadership evaluations of themselves and their colleagues. Results indicated that female officers were rated less positively than male officers. Further, non-white officers were rated less positively than Caucasian officers. Research by Boldry and colleagues (2001) with cadets showed that male and female cadets showed no differences in objective measures of performance. However, cadets' perceptions of their peers showed a pro-male bias, with male cadets being perceived as more motivated and possessing traits required of leadership positions. Female cadets were perceived as possessing more feminine traits, which were rated as likely to impair leadership performance. Consistent with this finding, research by Eagly and colleagues (1995) found that female military leaders were rated more positively if they were in a feminine leadership role than a masculine leadership role. The same was true for male leaders; specifically those in a masculine role (e.g., combat leaders) were rated more positively than those in a feminine leadership role (e.g., office managers).

Although researchers have documented stereotypes of women and ethnic minority leaders in military and non-military contexts, the effects of stereotype threat on women's and ethnic minorities' leadership performance remains unclear. Further, although women and ethnic minority officers are numeric minorities, it is not inevitable that they will experience stereotype threat. A decade of research on stereotype threat has clarified the parameters of effects on performance, yet we still know very little about the factors that lead some individuals who experience stereotype threat to psychologically disengage from these domains while others remain engaged and even excel in the face of negative societal perceptions.

The hypotheses guiding this research are: (H1) Participants in the stereotype threat group will have lower leadership performance and show a threat pattern in their physiological responses if they have strong identification with leadership, have strong gender identity (Study 1), or weak ethnic identity (Study 2); (H2) Participants in the control group will have adequate leadership performance and show normal physiological responses.

Research Approach

ROTC cadets from universities in southern California are recruited to participate in a study for women (Study 1) or ethnic minorities (Study 2). Participants are randomly assigned to either the stereotype threat (experimental) group or a control group. Before coming to the laboratory, participants complete an online questionnaire that measures personality characteristics that are expected to predict their leadership performance, including identification with leadership, gender identity (Study 1), ethnic identity (Study 2), resiliency (optimism, self-esteem, perceived control), anxiety, experiences with past discrimination, and expectations for future discrimination.

At the experiment session, participants complete another questionnaire that measures their expectations for their performance and their current thoughts and feelings about the leadership tasks. The experimental manipulation is then introduced by informing participants in the stereotype threat group that results from the leadership assessments they are about to complete will be examined for "group differences." This statement primes participants' social group status (race and gender) before they take the leadership assessments. Participants in the control group are told to think of the tasks as tests of creativity and there is no mention of group differences. By stating the tasks measure creativity, the tests are not viewed as diagnositc of leadership ability and should not arouse threat. Participants then complete the leadership assessment (Tacit Knowledge for Military Leaders) and play a simulated leadership in combat game (*Full Spectrum Warrior*). After the leadership assessments, participants complete a final questionnaire that reassesses their current thoughts and feelings about the leadership tasks. During key points in the study, including the pre-experiment, experiment manipulation, and post-experiment phases, participants' physiological activity will be monitored including their blood pressure, heart rate, and heart's electric activity (ECG or electrocardiogram).

Accomplishments

To-date, ROTC cadets have been recruited from Cal Poly Pomona, Claremont McKenna College, and through a mailing to ROTC cadets residing in the greater Pomona area. We are in the process of recruiting cadets from Cal State Fullerton and Harvey Mudd College. To bolster the sample, we added leadership-identified business students to the sample. Currently there are 22 participants, although data for 14 participants are reported here. There are 6 ROTC cadets, and 8 business majors. There are 6 females (2 ROTC cadets, 4 business students) and 8 males (4 ROTC cadets, 4 business students). Participants include 4 Asians (28.5%), 5 Caucasians (36%), 4 Latinos (28.5%), and 1 multiracial student (7%). There are 8 students in the threat condition (57%) and 6 students in the control condition (43%).

The multiple databases have been merged and processed. To-date the cardiovascular data have been scored, cleaned, and analyzed. This report describes the initial findings examining differences in cardiovascular reactivity between participants in the stereotype threat and control conditions, as well as differences in leadership performance.

The cardiovascular measures include systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), cardiac output (CO), ventricular contractility (VC), total peripheral resistance (TPR), and respiratory sinus arrhythmia (RSA). A threat response is characterized by increases in SBP, DBP, MAP, VC, TPR, no change in CO, and decreased RSA. A challenge response is characterized by little change in SBP, DBP, and MAP, increased CO, VC, and RSA, and decreased TPR. Reactivity scores were computed for each minute of the experiment by subtracting each minute of cardiovascular data from baseline. In all phases of the study participants in the experimental (threat) condition showed a threat response and participants in the control condition showed a challenge response (see Table 1.).

Table 1.
Mean Differences in Cardiovascular Reactivity by Condition

Phase	SBP, DBP, MAP		CO, VC		TPR, RSA	
	Exp	Con	Exp	Con	Exp	Con
Pre-Anticipation	SBP 11.75 > -2.64* DBP 13.11 > -8.26+ MAP 13.05 > -3.71**		CO *ns* VC *ns*		TPR *ns* RSA *ns*	
Leadership Test	SBP 9.67 > -10.68** DBP *ns* MAP 9.62 > -4.52**		CO -.525 < 2.11* VC 5.86 > -38.33+		TPR 1578 > -3457+ RSA -.96 < .545+	
Leadership Game	SBP 10.72 > -13.93** DBP 3.69 > -15.63+ MAP 24.68 > -9.69**		CO -1.31 < .928+ VC *ns*		TPR *ns* RSA *ns*	
Post-Appraisal	SBP *ns* DBP 2.74 > -4.99+ MAP 8.29 > -4.15*		CO -1.21 < .631** VC -6.86 < .001+		TPR 2524 > -2267+ RSA -.771 < .891^	
Recovery	SBP 5.30 > -9.97** DBP 8.23 > -19.51* MAP 8.66 > -1.64*		CO -.995 < 1.48^ VC -23.71< -1.80+		TPR 2359 > -3880+ RSA *ns*	

Notes. ** $p < .05$, * $p < .10$, + $p < .15$, ^ $p < .20$, *ns* = not significant. Values reflect mean differences from baseline.

There was a marginal interaction between condition and participants' systolic blood pressure (SBP) during minute 45 (the recovery phase) on leadership test performance, $F(1, 2) = 5.16$, $p = .15$. There were minimal differences in leadership performance by SBP among participants in the non-threat condition. However, among participants in the threat condition, those with high SBP during recovery had significantly lower leadership test performance than participants lower in SBP during recovery (see Figure 1). High SBP during recovery indicates a threat response whereas quick SBP recovery indicates a challenge response. Thus, participants in the threat condition who showed signs of physiological threat had lower leadership test performance. In contrast, participants in the threat condition who showed a physiological challenge response had high leadership test performance.

There was a marginal main effect of mean arterial blood pressure (MAP) on leadership game performance after controlling for gaming experience. Among participants in the threat condition, those with higher MAP during minute 26 (leadership game phase) of the experiment had lower leadership game performance than participants with lower MAP during minute 26, $F(1, 1) = 15.28$, $p = .159$.

The initial results examining leadership performance differences are marginally significant. However, with a sample of 14, this is to be expected. What is notable is that the cardiovascular data indicate clear threat and challenge responses even with a small sample size. It is expected that as the sample increases, additional effects will be detected. Further, additional analyses, such as the effects of personality variables as moderators (and buffers) of stereotype threat, will be feasible.

38
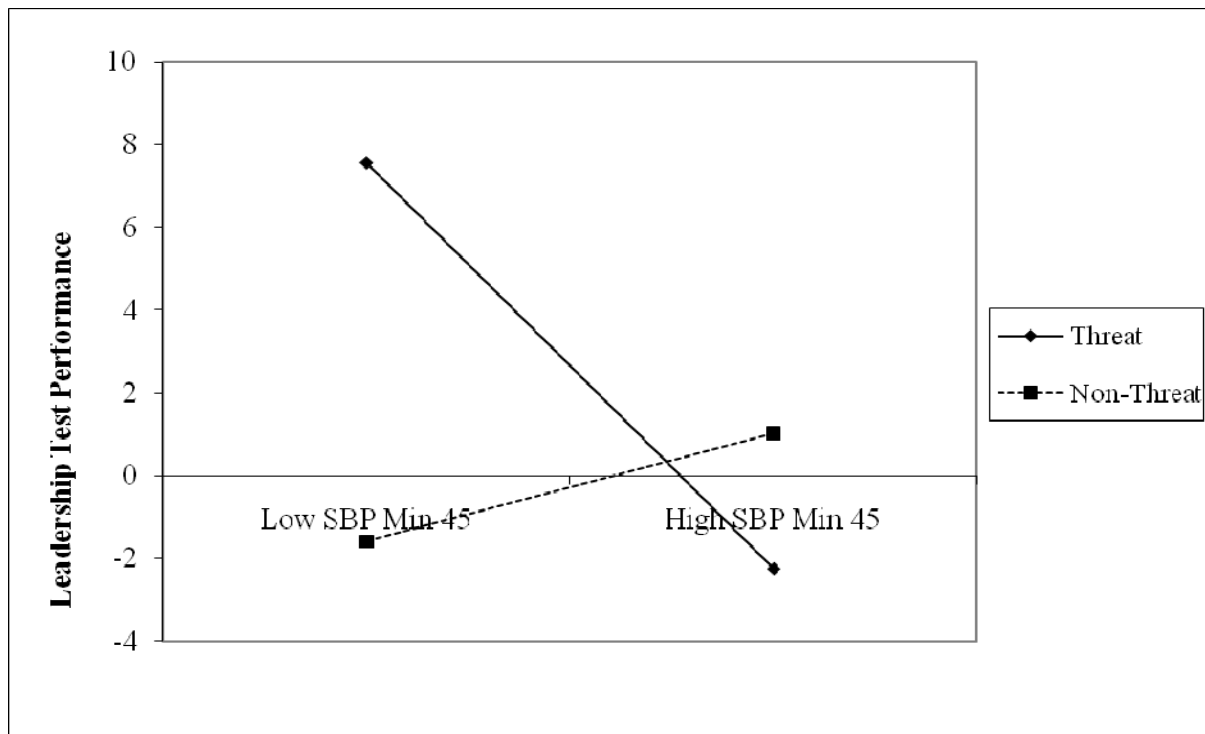
Figure 1. Interaction Between Condition and Systolic Blood Pressure on Leadership Test Performance

Contributions to Basic Science

This project will contribute to a more complex theory of the processes underlying stereotype threat. First, this study will show whether women and ethnic minorities in the military are negatively affected by stereotype threat about leadership performance. Second, findings will show which people are most likely to be affected, such as women who are highly identify with leadership and being female, or ethnic minorities who maintain ethnic pride and highly value leadership. The project examines the personality characteristic resiliency to see if people with this set of traits are less affected by stereotypes and negative expectations about their performance. Finally, this study will connect stereotype threat to physiological activity. People under stereotype threat are likely to show unhealthy cardiovascular responses and negative thoughts and feelings about their performance. People who have positive expectations should show healthy cardiovascular responses and positive thoughts and feelings about their leadership performance. Connecting thoughts, feelings, and physiological reactivity to actual leadership performance among ROTC cadets will provide insights into successful and unsuccessful adaptation to multicultural and high stress environments.

Potential Army/Military Applications

This research is consistent with the Army Research Institute's (ARI) overall goal of developing Soldiers who can effectively collaborate in quickly formed groups and in high stress

environments and developing Soldiers' leadership, interpersonal, and emotional skills. The research will provide insight into the conditions under which quickly formed and multicultural groups can work well together under stress. The personality characteristics that are found to predict resiliency to negative effects on leadership performance and physiological reactivity can be integrated into leadership development programs to further develop Soldiers' interpersonal skills and emotional control in multicultural, high stress environments.

Future Plans

Data collection will continue to the end of the contract period and beyond. It is anticipated that an ideal sample size will be reached by June 2010. The summer months (June through September) will be reserved for scoring the physiological data, conducting data analyses, and preparing reports, presentations, and a manuscript.

I plan to apply for continuation grants to carry out a follow-up study looking at whether mixed gender and racial groups suffer from intergroup anxiety and whether this anxiety negatively affects collaborative leadership performance.

References

Biernat, M., Crandall, C. S., Young, L. V., Kobrynowicz, D., & Halpin, S. M. (1998). All that you can be: Stereotyping of self and others in a military context. *Journal of Personality and Social Psychology, 75,* 301-317.

Blascovich, J., Mendes, W. B., Hunter, S. B., & Lickel, B. (2000). Stigma, threat, and social interactions. In T. F. Heatherton, R. E. Kleck, M. R. Hebl, & J. G. Hull (Eds.). *The social psychology of stigma* (pp. 307-333). New York: Guilford Press.

Boldry, J., Wood, W., & Kashy, D. A. (2001). Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues, 57,* 689-705.

Boyce, L. A., & Herd, A. M. (2003). The relationship between gender role stereotypes and requisite military leadership characteristics. *Sex Roles, 49,* 365-378.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. *Psychological Science, 16,* 572-578.

Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology, 88,* 276-287.

Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 680-690.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*, 573-598.

Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychological Bulletin, 117,* 125-145.

Mettrick, J., & Cowan, G. (1996). Gender stereotypes and predictions of performance: Women in air combat. *Journal of Social Behavior and Personality, 11,* 105-120.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52, 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual performance of African Americans. Journal of Personality and Social Psychology, 69, 797-811.

Examining Leadership Processes in Complex Network Environments

Contract #: W91WAW-08-C-0028
Institution: University of Central Florida

Contract Dates: 03/01/2008 to 07/31/2010
PI: Leslie DeChurch
Co-PI: C. Shawn Burke

Problem(s)/Research Question(s). What are the essential leadership processes that enable networks of teams to operate effectively? In particular, are top-down or laterally-implemented leadership functions more impactful to network functioning?, are leaders more impactful when they intervene during teams' planning (i.e., coordination by plan) or action/performance phases (i.e., coordination by mutual adjustment)?, and what are the cognitive and behavioral mechanisms through which various enactments of leadership impact network functioning?

Technical Barrier(s). Models of team performance in networked environments are not well developed, nor are models of leadership that link leadership to core aspects of performance in networked-team environments.

Significance/Impact for Basic Research. Understanding how level of enactment of leadership (top down v. laterally) and phase of team performance in which leadership intervenes (transition v. action) impacts the synchronization of both individual teams and networked teams will improve models of leadership, networked team performance, and the conceptual model of leadership and performance in networked-team environments developed in this research effort. Furthermore, this research examines the effects of having multiple leaders operating at different levels of the system simultaneously.

Potential Transitions. Knowledge gained in this basic research effort may be useful for:
- ARI's leadership development research programs.
- ARI's applied training research programs.
- Any research applying network science to organizations

Overview

Collaborative work systems offer the ability to address complex problems by combining expertise distributed across business functions, knowledge specialties, and geographic locations. Often, the goals these systems face are complex and multifaceted requiring multiple distinct teams to coordinate their efforts with other distinct teams as an integrated network of teams. Because of the inherent diversity of membership and complexity of goals, leaders play a vital role in ensuring network-level success. However, there is an increasing realization that many theories of leadership do not address the unique leadership requirements of teams, particularly in a multiteam system (MTS) environment.

The primary question addressed by this research is: What are the essential leadership processes that enable networks of teams to operate effectively? We approach this problem by examining the functions of leaders that direct individual component teams and their members, and also the functions of leaders directing an overall system or network of teams. This research is paramount in elucidating similarities and differences between team leadership and leadership in the context of a complex network of teams. Figure 1 presents and overview of the theoretical

model we test which links leadership to teamwork both within and across teams, and then to overall system performance.
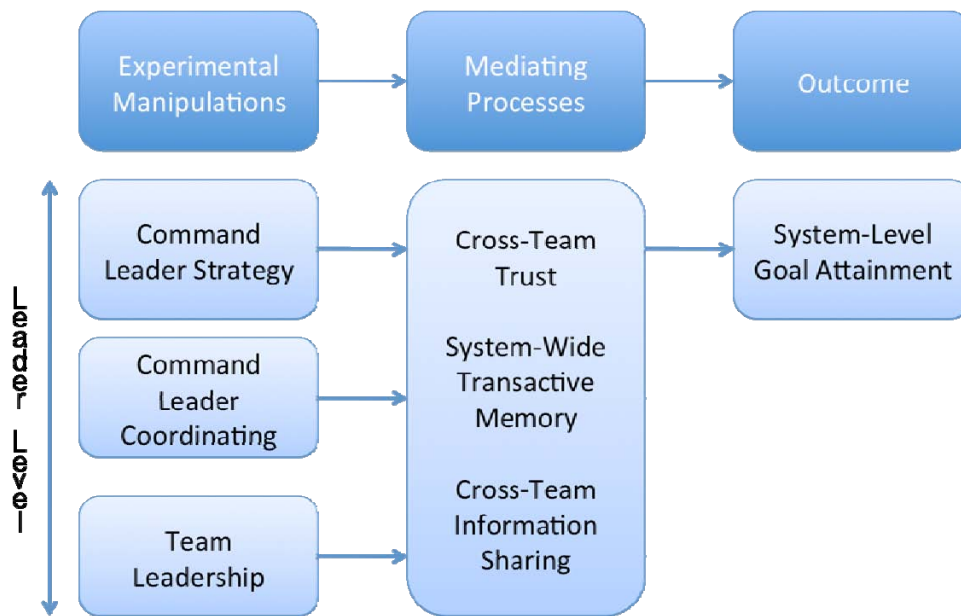
Figure 1. Summary of relationships between leadership and multiteam performance examined in Year 1.

Understanding the leadership processes and skills needed to effectively direct and coordinate effort in a multiteam performance environment is particularly important for the U.S. Army, where MTSs are commonly used in stability and support operations, as well as combat missions. Though the task tempos of these two types of situations differ greatly, both require multiple specialized teams to synchronize their efforts to interface effectively, which necessitates that leaders must coordinate efforts both within and across teams.

Research Approach

To understand the role of leaders in networks of teams, we designed a laboratory simulation task where we could manipulate various combinations of leadership in a team-of-teams arrangement that mirrored key features of Army teams. Two 3-person teams interacted virtually with each other, and with 17 simulated AI-controlled teams to perform a humanitarian aid mission. Teams were led by both an internal team leader and a confederate system-level commander. The system goal was to enable a convoy of humanitarian aid to safely pass through a hostile region. The two live teams worked on markedly different functional goals (route clearance, insurgent removal), facing competing pressures to work towards their team goal versus the larger humanitarian goal.

We conducted an experiment using 432 undergraduate students arranged into 72 multiteam systems. Each was tested in a separate 5-hour session. Findings reveal command level leadership that occurs during team task accomplishment was most pivotal to accomplishing the larger

humanitarian goal. This form of leadership benefited the system in two key ways: giving rise to a multi-team transactive memory system whereby teams had a heightened awareness of which teams possessed which expertise, and improving the extent to which distinct teams planned together and shared information openly. Command strategy which occurs in advance of the task was not an effective form of leadership. Laterally-enacted (team) leadership was effective, by improving cross-team trust, giving rise to cross-team transactive memory, and aiding in cross-team information sharing. Unexpectedly, we found leadership enacted from multiple levels simultaneously actually impeded the functioning of the system. Command leader strategy was particularly harmful, as it led teams to be overly rigid and not-responsive to the guidance of team leaders. This pattern is illustrated in Figure 3.
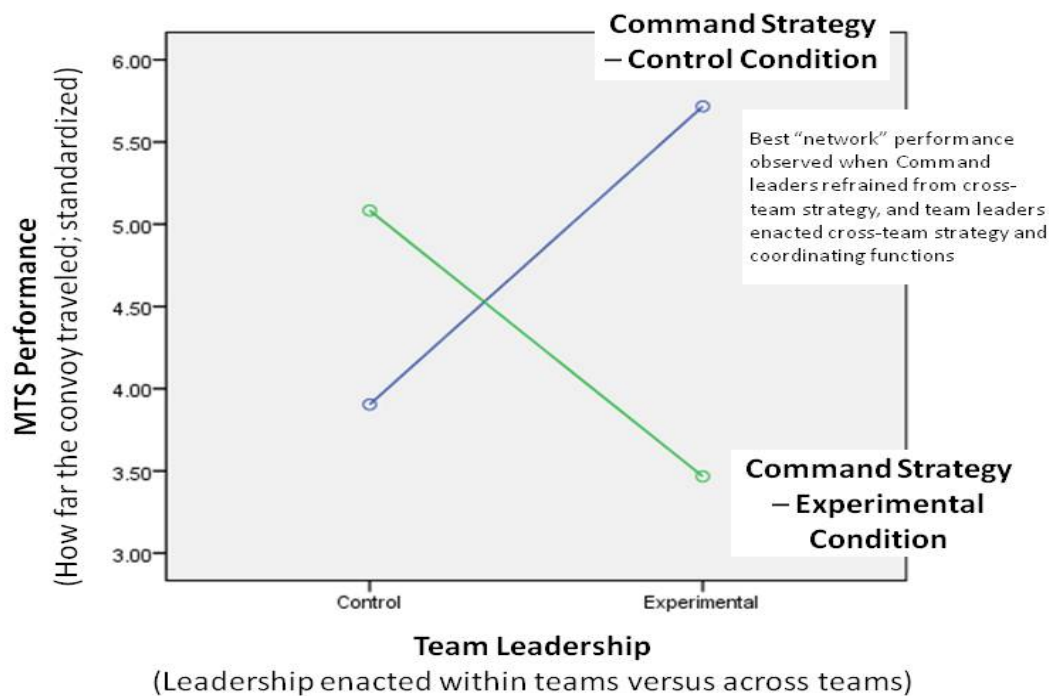
Figure 3: Neutralizing interaction between team and command level leadership (Year 1).

The current effort overcame technical barriers by first constructing a theoretically grounded systems model of network performance that specified the role of leadership and emergent characteristics, and then experimentally testing core propositions about leadership derived from this model. As research on leadership in complex network environments is in its infancy, a premium was placed on the ability to draw firm causal conclusions about the impact of leadership on various aspects of network functioning. The overall approach of this research advanced both models of team performance in networked environments and models of leadership that link leadership to core aspects of performance in networked-team environments.

Accomplishments

During the first half of this year, we completed data collection and analysis for the first year, and began dissemination efforts. Due to the rich data collected and broad theoretical impact for organizational science, we have been widely disseminating year 1 findings. The PI (DeChurch) has given 3 invited talks on this research (Drexel University, University of South Florida, and Grenoble Ecole de Management), and presented 3 conference presentations based on our findings (Society for Industrial/Organizational Psychology, INGRroup, Academy of Management Meetings). A research scientist on the project (Doty) presented our simulation design in a multiteam system symposium at the SIOP Meeting. Project personnel have co-authored a leadership review based on this work, which was invited for revision at the *Journal of Management*, and a book chapter currently in press. We are continuing dissemination efforts with additional presentations at research conferences and journal submissions.

Although our year 1 findings provide meaningful insights on the level of enactment of leadership, and functional leadership needs, these findings raise two important issues which are now the focus of a second (Year 2) experiment. The first issue: if leadership needs to be enacted during task accomplishment, without functional redundancy from multiple leaders in the system, how can leadership best be distributed within the system to optimize performance? To address this issue, we are now manipulating the configuration of leadership and comparing systems where leadership is fully shared (control) to those where leadership is distributed within the network (experimental). The second issue arose out of the practical reality that often these systems fail to function because they lack the basic building blocks of an intact social system: task and socio-emotional. In multiteam systems the task breakdown faced stems from the fact that teams are typically distributed and face difficulties in communicating. The largest socio-emotional barrier is a lack of trust across teams. Thus, we have created two additional manipulations: trust and communication reliability, in order to examine the direct and interactive effects of leadership configurations and task and socio-emotional capacity.

In year 2, we will address the following questions: Which configuration of leadership is most beneficial to system functioning? In what ways do (a) trust breaches and (b) communication breakdowns undermine the performance of teams-of teams? Which type of breach: socio-emotional versus task, are systems most susceptible to? Which configuration of leadership best creates resiliency in multiteam systems facing trust and communication breakdowns? In addition to examining these important substantive issues, we have completed a simulation reconfiguration that enables us to examine multiple layers of collective identity in more complex distributed virtual team arrangements. Figure 4 presents and overview of the relationships being examined in the Year 2 experimentation.
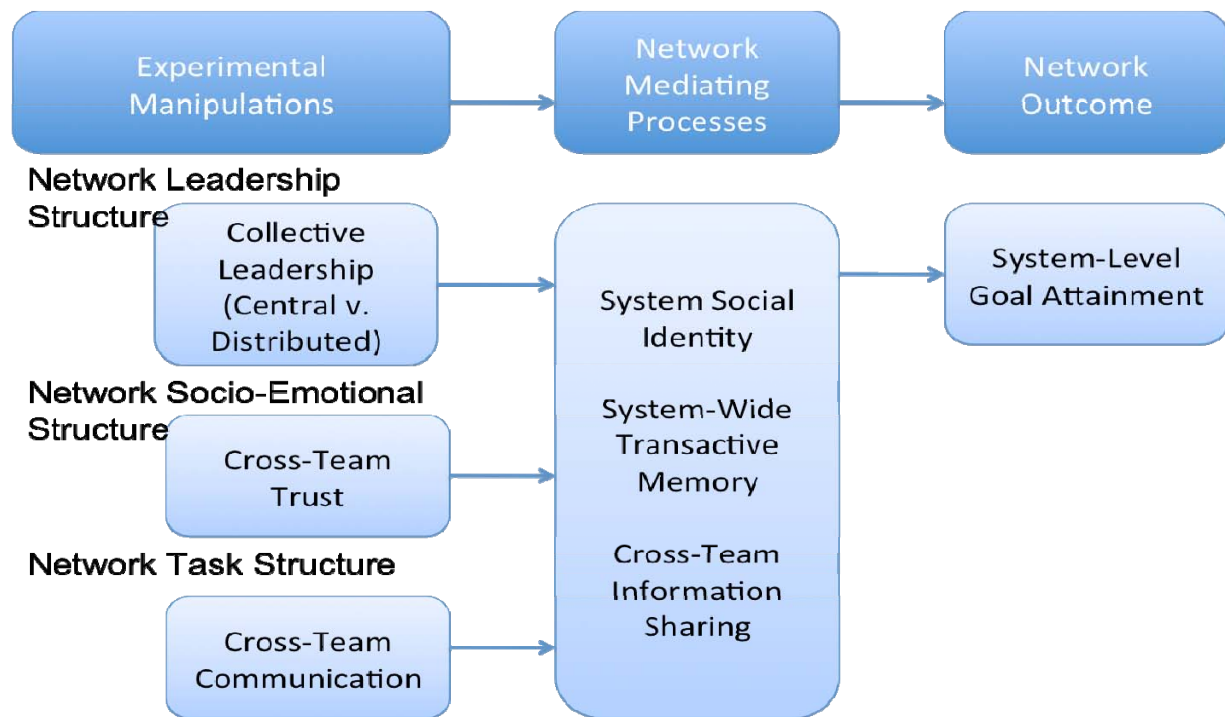
Figure 4. Summary of relationships between collective leadership, trust, and communication reliability and multiteam performance examined in Year 2.

Contributions to Basic Science

Thus far, this effort makes two significant contributions to basic science. First, this effort extends thinking about leadership to consider the joint effects of leadership being enacted from multiple levels. Interestingly, leadership enacted at one level is an important contextual factor affecting the utility of leadership enacted from another level. Findings show that leaders can have a positive impact on the functioning of the system from either the command level or the team level, but that leadership is best enacted from one level or the other. Further, networks of teams function best when leadership maintains alignment during task engagement, and not in advance through planning.

The second contribution is to organization science. In organizations such as the US Army where many teams are simultaneously engaged in large-scale efforts, there are tightly coupled constellations whose efforts may be individually successful to the detriment of the larger mission. Current findings show the factors that make teams successful (e.g., trust, shared mental models) can actually harm the ability of teams to work together across their boundaries. This effort demonstrates that leaders can be instrumental in overcoming this within-team bias to the extent that they target their actions across teams—at the point of team-to-team interfaces—rather than downward, within their respective units.

47

Potential Army/Military Applications

Some of the anticipated products of this research effort are in the areas of training design and leadership development. Current findings which link aspects of leadership to important characteristics and outcomes of networks of teams will enable the development of training programs and feedback and appraisal systems that support the development of these critical functional behaviors in Army leaders. In particular, the current research effort will provide guidance for the impact of level of leadership and timing of intervention in the task cycle that will inform the design of future applied leader development efforts. In addition, uncovering the unique contributions of a variety of characteristics of teams and networks of teams to overall network performance will enable the development of training programs for Soldiers operating in networked-environments that specifically promotes skills necessary for effective interaction within a network.

Future Plans

We are currently collecting pilot data for a second experiment; in the remaining months of this second year we will:
- Complete data collection of the laboratory experiment manipulating 3 multiteam system characteristics: cross-team trust, collective leadership, and communication reliability.
- Complete data analysis & write a report summarizing the major findings of year 2's research effort.
- Prepare and present additional papers on the contributions to organizational science learned in year 1. (dissemination effort).
- Prepare and present papers on year 2 findings. (dissemination effort).

Language and Social Dynamics

Contract #: W91WAW-07-C-0029   **Contract Dates:** 05/21/2007 to 05/20/2009
Institution: University of Texas at Austin   **PI:** James W. Pennebaker

Problem(s)/Research Question(s) – What can we tell about the social dynamics of a working group based on the ways group members talk with one another?

Technical Barrier(s) – The measurement and interpretation of natural language in ongoing social groups is a new research focus. Linking group dynamics to the use of function words (e.g., pronouns, prepositions, and related words) has only recently become possible with the advancement of computerized text analytic tools.

Significance/Impact for Basic Research – How people use function words in everyday conversation provides information about how they are thinking and relating to other people. The analysis of language cues that people cannot readily manipulate (i.e., function words) provides a non-reactive and unobtrusive way to measure group dynamics.

Potential Transitions – Knowledge gained in this basic research may be usefully employed in:
- ARI and DoD management and group training programs,
- the Intelligence Community's interest in remote sensing of groups and group dynamics, and
- The U.S. military and intelligence efforts at cultural interpretations.

Overview

Two relatively new overlapping technologies allow us to study language in groups: computerized text analysis programs and the Internet-based world of social communication. Historically, the analysis of language has focused on the *content* of what is being conveyed. Now, it is possible to explore the linguistic *style* that surrounds that content: we are focusing on people's use of function words, such as articles, auxiliary verbs, prepositions, and pronouns.

Function words are processed in the brain very differently than content words and are uniquely tied to basic social and psychological processes. We have linked the use of various function words with demographic factors (e.g., age, gender, social class), personality traits (e.g., self-esteem, neuroticism, conscientiousness), social factors (e.g., dominance, honesty, marital satisfaction, testosterone levels), and psychological states (e.g., depression, suicide-proneness).

We have started to explore three new areas of research with groups that interact face-to-face or by computer-mediated communication. The first broad research question asks how linguistic style can be used as a measure of social climate, including group task performance, liking, and leadership. The second research question examines linguistic style of leaders and effective leaders. What language cues predict who will emerge to be a leader in an initially leaderless group? What are the language styles of effective leaders? Our third area focuses on the development of text analytic tools and data mining.

Research Approach

Our research uses a relatively non-intrusive methodology to assess individuals and groups in terms of cohesiveness, functioning, and individual-group relationships. Our primary

program, Linguistic Inquiry and Word Count (LIWC), computes the percentage of words in a text that belong to a host of linguistic (e.g., articles, pronouns, conjunctions) and psychological categories (e.g., positive and negative emotions, cognitive mechanism words, social words). We are studying function word use in two ways. The first way is to simply compare the actual rates of function word use by different people in a group or by different groups as a unit. For example, we can determine if successful groups use language differently from unsuccessful ones. The second way is to examine how each person in a given group matches function word rates with the other members of the group. This language synchrony measure is referred to as Language Style Matching, or LSM. Mathematically, LSM is based on how often a person uses each of nine function word categories relative to the average of their group. By isolating people's ability to synchronize with others, LSM provides a metric of social competence in a complex environment.

Beyond our lab experiments on group task performance, we are analyzing group dynamics using written and transcribed oral transcripts. Part of this research focuses on large natural language communities (e.g., emails, blogs, and wikipedia) where we also have access to measures of social cohesiveness, relative status, or conflict. By measuring word usage and LSM, we are evaluating the degree to which groups work closely with one another and exhibit signs of tension or productivity. Lastly, we are developing additional tools and algorithms to find meaningful patterns in natural language use, and for data mining natural language samples on the Internet. Part of this research is building our text analysis tools for other languages, such as Arabic.

Accomplishments

Language Style Matching in Groups: Our previous lab work showed that while verbal synchrony is related to attraction in dyads and small work groups, it does not indicate that groups are working more effectively in a cooperative task. Rather, LSM is a measure of when people are trying harder to "get on the same page", or to establish common ground.

Our more recent series of lab studies of LSM in simulated negotiations provide further evidence that LSM is a marker of engagement and shared attention (Ireland & Henderson, under review). In two studies, LSM was highest in competitive negotiations between dyads where no agreement was reached (see Figure 1). In addition, LSM was associated with less efficiency (as measured by using more words and more time taken) in those cases where an agreement was reached. A third lab study showed that ratings of contentiousness in the interaction fully mediated the effects of LSM and likelihood of reaching an agreement. These studies show that rather than only reflecting positive engagement, LSM reflects engagement or shared attention more generally. In the case of competitive interactions, then, LSM can indicate conflict or negative outcomes.

Figure 1. LSM in Negotiations where an Agreement is Reached or Not
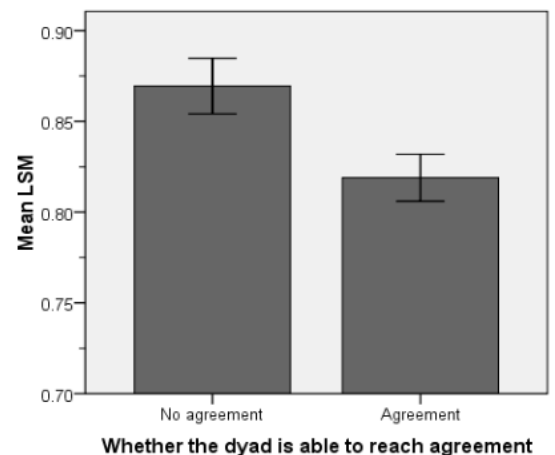
Note. No agreement indicates that dyads failed to reach a mutually satisfactory agreement on all issues on the table within the allotted time. Error bars are 95% confidence intervals.

To assess the ability of LSM to reflect engagement within relationships over time, we examined archives of written letters and professional work between well-known pairs of spouses and friends over the course of their relationship (Ireland & Pennebaker, under review). The pairs included Sigmund Freud and Carl Jung, Robert and Elizabeth Barrett Browning, and Sylvia Plath and Ted Hughes. LSM was highest during times of relationship harmony and lowest during times of illness and relationship disharmony. Importantly, these studies showed that LSM changes reliably in response to relationship functioning over time in naturalistic contexts.

We are now finding that other metrics of language style mimicry can be used in investigations of author identification (Ireland & Pennebaker, in preparation). We have found individual differences in the degrees to which playwrights and screenwriters are able to accurately mimic the linguistic style of the opposite sex in written dialogue. This is the first in a series of studies testing the ability of computerized text analyses to determine the degree to which authors write or speak like members of another group or subculture as a function of familiarity, contact, or identification with another group. Note that a linguistic style approach goes beyond obvious key terms (specialized topics, slang) and common syntactic constructions that characterize group membership and that are readily manipulated. This uniquely places the assessment of linguistic style as a potential method for remote sensing in issues of group identification, defectors, or imposters.

Language and Status. Across many lab studies, we have found that higher status participants used more words, fewer "I", and more "we" than lower status participants (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, under review). In addition, LSM is higher when participants were more equal in status. That is, the more similar the ways people use function words, the smaller the difference in status between them (Kacewicz, Pennebaker, Davis, Jeon, & Graesser, unpublished data). Furthermore, we consistently find that interactants in dyads and within leaderless groups (of 4-6 people) highly agree on who has high status within the group (Kacewicz, Beer, Srivastava, & Pennebaker, in preparation). However, when there are discrepancies in an individual's self-rated status and members'-ratings of the target's status, this self-enhancement (i.e. when an individual thinks they are higher status than they are actually perceived) is reflected in language through more "you" use, more negative emotion words, and fewer assents (e.g., absolutely, agree, yeah).

Our latest studies on a corpus of IMs exchanged over 15 months between 22 employees of a research and development firm confirmed and built upon our previous findings (Scholand, Tauszik, & Pennebaker, 2010). The relative rates of first person singular pronoun use by the employees were used to accurately map within the working group. "I" was used more by those in lower status roles. Individuals who were rated by the employees as having higher status used more first person plural pronouns, and tended to style match less with their interaction partners on average, (i.e. they had lower average LSM scores; see Figure 2; Tauszik, 2009). This might reflect lower status individuals' attempts to work harder to get on the same page as their higher status counterparts.

Figure 2. LSM as a Function of Rated Social Status
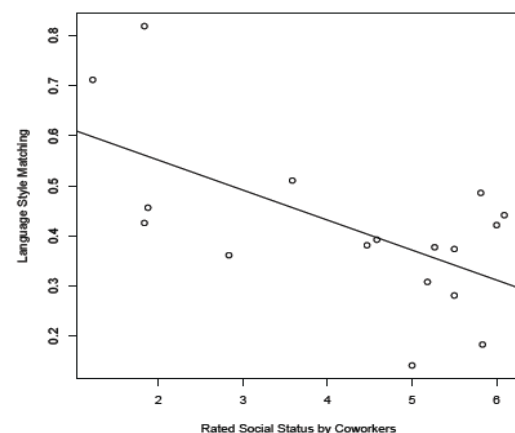
Note. Correlation of mean rated social status by coworkers with mean LSM. r(15) = -0.62, p < 0.01.

The same linguistic style markers of status in English were found in a sample of 60 translations of memos between members of Saddam Hussain's administration (Hancock, Beaver, Chung, Frazee, Pennebaker, Graesser, & Cai, under review). Specifically, higher status individuals used significantly fewer words, marginally fewer first person singular pronouns, and significantly more second person pronouns. This study is particularly relevant to the Army Research Institute because it shows that our assessments of status based on natural language are valid in military communications.

Language in Communities. The examination of larger groups can help us to identify and understand patterns in how relationships form, how group identity and commitment develop, and how emotional reactions and information spread in a network. For example, 9,508 blogs that mentioned swine flu were downloaded during the period from when the World Health Organization had publicly announced the appearance of swine flu, April 24, 2009 to May 7, 2009. The number of blog entries mentioning swine flu showed a curvilinear pattern, and was highest during the period of April 29 to May 1st. Using LIWC's emotion categories, it was found that perceived anxiety levels were highest early on, despite the fact that the actual risk of infection was greater later. The initial jump in anxiety may have been warded off by publicized (low) morbidity rates, and replaced by anger, which is consistent with research that has reported increases in intergroup hostility when an infectious disease originates in a foreign country. These emotional responses are consistent with the literature on reactions to emerging infectious diseases, but are unavailable for real-time or near real-time measurement using traditional methods.

Another study looked at word use and social networks in an online blog community devoted to weight loss (Chung, 2009). Having larger weight loss goals and blogging about personal events (i.e. using more first person singular pronouns) was a more effective weight loss strategy than keeping an online food intake diary. The degree to which bloggers were socially integrated with the blog community (i.e. commented to many other bloggers and received many comments) was found to be a potent predictor of weight loss. The results showed that the size of one's social network is a significant predictor of goal success within an online community.

Development of Text Analytic Tools. We are now completing an Arabic text analysis program of function words based on our current computer program (named LIWC) that uses an English categorization scheme that would read Arabic text (Hayeri, Chung, & Pennebaker, 2009-a). For the validation phases of the dictionaries' development, we have compiled one of the largest corpuses of Modern Standard Arabic texts. We hope to give people who work with translators and translations an insight into dimensions of a culture that may be invisible to someone unfamiliar with the other language or culture.

Some of our preliminary work with the Arabic LIWC is showing gender differences in the use of high (formal) Arabic in texts (Hayeri, Chung, & Pennebaker, 2009-b). We compared language use by Maryam Jameelah and Abi-l-A'la al-Maudoodi in a series of written letters between the two scholars between the years 1960 and 1962. We also compared the language style in speeches delivered by Hosni and Suzanne Mubarak during the years 2002 and 2003. We found that the women in these pairs tended to adopt a linguistic style closer to the Qur'an (the "high" variety of Arabic) than did men. Further analyses are being conducted to see whether these differences in formal language use hold up across translations.

Our work in the development of text analytic tools is underway for Russian, Turkish, and Chinese with the help of experts around the world. A Persian dictionary is our newest project in this vein, headed by language experts from the University of Tehran.

Contributions to Basic Science

Language is generally thought of as a tool that we consciously manipulate. Changes in one's pitch or rate of speech happen outside of conscious awareness, but language is generally assumed to be a highly intentional communication channel. Our findings suggest that even the words we use are highly susceptible to influence by the affective nature of social dynamics; certain linguistic categories, such as function words, are as much a sign of how we feel as they are a tool for expressing it. The analysis of function words provides important clues about the social relationship between a speaker and the audience, the psychological state of the speaker, and information about the speaker's social role and status in the community.

Through the development of our computerized text analytic tools, beginning with Arabic, we hope to uncover the function words that reflect social dynamics and psychological states across cultures in translated texts. An examination of the function words that are dropped or added in a translation can provide explicit, novel, and measurable information to a culture's interpretation of a text.

Potential Army/Military Applications

Language analysis can be viewed as a form of remote sensing. Whereas we have always known that what people are saying is important, the primary contribution of our work is in noting that how they are conveying information also provides important insights. We are beginning to see reliable language patterns across various experiments that we have conducted using small groups in tasks. Overall, our results suggest there is a particular linguistic style associated with being a leader and with being a lower status individual. These findings provide a baseline for identifying leaders and serve as a foundation for identifying the language of effective leaders, and engagement of group members.

The military makes significant decisions based on the translations of interviews, speeches, documents, and other verbal records. With the development of comparable Arabic text analytic tools, we can provide translators, interpreters, or end users of translations with information about the social and psychological dimensions of any type of text sample. The underlying technology is transparent and the final computer programs are fast and easy to run. This project is both theoretically compelling and has an immediate, real world application.

Future Plans

Our plans for the next year of funding are to apply the findings from our lab to real world communities. Given the amount of language and other information available in online communities, these corpora continue to serve as vast resources for piecing together a bigger picture of how groups function. We are continuing to assess what predicts the size of a social network and identifying the "influencers" in a network based on language use within the diet blog community. Finally, we have downloaded all texts from Wikipedia, totalling over 2 terabytes of interactive and longitudinal group data. From this corpus, we hope to assess how information becomes socially agreed upon, and the ways in which entries are edited and deleted based on domain expertise as well as reputation and experience within Wikipedia's forums. For our cross-language efforts, we have prepared a corpus of political speeches and interviews in

both English and Arabic to pilot our Arabic LIWC software. While our research aims to uncover group dynamics through language use, we are continually developing our data mining and text analysis toolkits that may be used for basic and applied research across the small groups domain.

References

Chung, C. K. (2009). *Predicting weight loss in blogs using computerized text analysis.* Unpublished dissertation. Austin, TX: The University of Texas at Austin.

Chung, C. K. & Pennebaker, J.W. (2007). The psychological function of function words. In K. Fiedler (Ed.), *Social communication: Frontiers of social psychology* (pp 343-359). New York: Psychology Press.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2008). Language style matching as a predictor of social dynamics in social groups. *Unpublished manuscript.*

Hancock, J. T., Beaver, D. I., Chung, C. K., Frazee, J., Pennebaker, J. W., Graesser, A. C., & Cai, Z. (accepted with minor revisions). Social language processing: A framework for analyzing the communication of terrorists and authoritarian regimes.

Hayeri, N., Chung, C. K., & Pennebaker, J. W. (in preparation-a). Arabic Linguistic Inquiry and Word Count: Viewing the world through English and Arabic eyes.

Hayeri, N., Chung, C. K., & Pennebaker, J. W. (in preparation-b). Gender differences in High Arabic using Arabic Linguistic Inquiry and Word Count.

Ireland, M. E. & Pennebaker, J. W. (under review). Language style matching in reading and writing. Manuscript submitted for publication.

Ireland, M. E., & Pennebaker, J. W. (in preparation). Gender differences in the verisimilitude of fictional dialogue.

Kacewicz, E., Beer, J., Srivastava, S., & Pennebaker, J. W. (in preparation). The language of status self-enhancers and status self-effacers in leaderless group discussions.

Kacewicz, E., Pennebaker, J. W., Davis, D., Jeon, M., & Graesser, A. C. (under review). Pronoun use reflects standings in social heirarchies. Manuscript submitted for publication.

Niederhoffer, K.G. & Pennebaker, J.W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21,* 337-360.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Linguistic Inquiry and Word Count (LIWC2007): A text analysis program.* Austin, TX: LIWC.net.

Scholand, A. J., Tausczik, Y. R., & Pennebaker, J. W. (2010). Social language network analysis. To be presented at the xyz.

Tausczik, Y. R. (2009). *Linguistic analysis of workplace computer-mediated communication.* Unpublished Master's thesis. Austin, TX: The University of Texas at Austin.

Tausczik, Y. R., & Pennebaker, J. W. (2009). The effects of distraction in language style matching over IMs and phone calls. Unpublished data.

Tausczik, Y. R., & Pennebaker, J. W. (in press). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology.*

BRU RESEARCH OBJECTIVE #3: PROVIDE FUNDAMENTAL KNOWLEDGE FOR IDENTIFYING, ASSESSING, AND ASSIGNING QUALITY PERSONNEL FOR THE EVOLVING ARMY

Research in this section is directed toward identifying and measuring the aptitudes and skills that are unique to the human performance requirements of the Future Force. Exploring the sociological and psychological factors that could influence recruitment, retention, and Army performance are important aspects of this research objective.

The Link Between Feeling and Seeing

Contract #: W91WAW-08-C-0018   **Contract Dates:** 06/01/2008 to 05/31/2011
Institution: Boston College   **PI:** Lisa Feldman Barrett

Problem(s)/Research Question(s)

What you see influences how you feel. But is the opposite true? Does feeling influence vision?
Does a person's internal state (experienced as pleasant or unpleasant with some degree of
arousal) called "affect" influence vision? Specifically, we are testing whether feeling influences
what people see (called visual awareness) and their ability to detect and act on visual information
in the absence of subjective awareness (called objective awareness or blindsight). We call this
the "seeing with feeling" hypothesis.

Technical Barrier(s)

This is a novel hypothesis that derives from the neuroanatomical connections between parts of
the brain involved in representing bodily feelings and those involved in vision.

Significance/Impact for Basic Research

Our commonsense experience tells us that we first see objects in the world, and then we evaluate
them for the relevance or value – for their significance to us at the present moment. First we see
a Macintosh apple, and then we decide if we are hungry or if we'd prefer a Red Delicious. First
we see a scowling face and then we decide whether or not it is threatening, and if we like the
person. Many theoretical models of emotion and of vision assume this is the state of affairs.
Based on careful neuroanatomical study, however, we instead hypothesize that how a person
feels at the time of perception will influence what is actually seen. External visual sensations do
not occur in a vacuum, but within a context of internal sensations from the body that houses the
brain. Instead of directly experiencing changes in their breathing, muscle tension, or stomach
motility, people routinely experience more diffuse feelings of pleasure or discomfort, feeling
worked up or slowed down. In this research, we explore the hypothesis that the brain routinely
takes bodily information into account during vision. From this perspective, sensations from the
body are a dimension of knowledge – they literally help us to identify what the object is when we
encounter it, and how to act on the object, based in part on past reactions. If the "seeing with
feeling" hypothesis is correct, then affective responses signaling an object's salience, relevance,
or value, do not occur as a separate step after the object is identified – a person's affective
response assists them in seeing an object from the very moment that visual stimulation begins
(Barrett & Bar, 2009). The implication is that people don't come to know the world exclusively
through their external senses; rather, their affective states influence the processing of sensory
stimulation from the very moment an object is encountered. These ideas suggest the intriguing
possibility that exposure to visual sensations alone is not sufficient for vision to occur. Taken
even further, the "seeing with feeling" hypothesis suggests other several intriguing possibilities.
It may be that individuals who experience intense affective responses may literally see the world
differently (it's not just a matter of interpretation). Furthermore, it is highly possible that certain
types of life experience (e.g., prolonged drug use, living in urban poverty or with chronic stress)

can wire people to see things that precipitate psychological distress, making them more susceptible to mental and physical illness.   When taken even further, these ideas suggest that exposure to visual sensation alone is not sufficient for conscious "seeing" – those sensations have to be experienced as personally meaningful.

Potential Transitions

Knowledge gained with this basic research might be usefully employed in:
- ARI's applied training research programs
- DOD or Homeland Security research on how to enhance a perceiver's ability to detect a threat.

Overview

When an American Soldier enters a house in an Iraqi suburb, or an Afghan neighborhood, he or she must quickly decide whether he is facing a friend or a foe. This decision will have consequences for somebody's life. We are interested in the following questions: Does the Soldier's feelings of anger, fear, or even excitement influence his or her ability to detect or literally see an Iraqi or Afghan national scowl or cringe or even perhaps smile? Would this task be made harder or easier by manipulating the surroundings and external conditions?   The current research asks the very basic question of how the state of the perceiver (which is referred to as "affect") influences what is detected and consciously seen.

Research Approach

We are investigating two different hypotheses: First, do feelings influence what is literally seen? If you are angry or afraid (in an unpleasant, high arousal affective state), will your brain choose to consciously focus on the face of another person, or perhaps something more innocuous, like a house? To test this hypothesis, we use a paradigm called "binocular rivalry." Binocular rivalry occurs when incompatible images are presented to each eye (e.g., a face to one eye and a house to the other) such that the two cannot be merged into a coherent 3-dimensional image. Instead of perceiving a mixture of the two images, people experience one image at a time, oscillating back and forth into visual awareness every few seconds. By measuring which percept is seen first, and for how long, it is possible to assess what visual information is selected for subjective visual awareness. We are currently exploring how changing the affective state of the perceiver more directly influences what the brain chooses to see.

Secondly, we are investigating whether feelings enhance or diminish the ability to detect a threat outside of consciousness (called objective awareness or functional blindsight). If you are angry or afraid (in an unpleasant, high arousal affective state), will it be easier or harder to detect a threat as it flashes before you, even if you can't consciously see anything? Real blindsight occurs when perceptually blind people (i.e., people who report not being able to see) are able to detect and act on visual stimuli without having any awareness that they can do so. Blindsight can be induced in the lab with the brief presentation of an object (e.g., 16, 33, or 67 ms) followed by a backward mask (to prevent feedback processing that is necessary for subjective awareness). Although objects are not consciously seen under these conditions, people can sometimes still

respond to them behaviorally in such a way as to indicate that the objects are being detected at better than chance levels. We hypothesize that a strong affective state may enhance experimentally-induced blindsight, so that intense feelings may allow people to better detect and act on certain objects or blind them to others, before the sensory information is shaped into a fully formed percept that reaches subjective awareness. Specifically, we are using a modified signal detection paradigm to test the blindsight hypothesis.

Accomplishments

The Influence of Feelings on Visual Experience

In Experiment 1, we asked whether changing a person's feelings in response to a neutral object would impact whether that object is seen (Anderson, Bliss-Moreau, & Barrett, under review). Participants learned that neutral faces were either "good", "bad", or "neutral" by pairing them with descriptions of positive, negative, or neutral descriptions of behavior ("helped an old woman cross the street", "kicked a dog", "mailed a letter"). Participants then completed a binocular rivalry task in which the faces alone were presented with neutral objects (houses) (Figure 1).

Left eye   Right eye
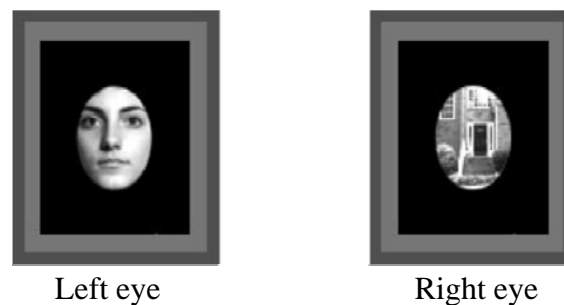
Figure 1. Binocular rivalry images

Although all the faces were structurally neutral, those that had been paired with negative behavioral descriptions were seen for longer (i.e., dominated in visual awareness) than those paired with neutral or positive descriptions (Figure 2). This means that the brain is prioritizing visual representations with negative affective value. These findings illustrate that affective information can influence vision in a completely top-down way that is not attributable to the visual features of the objects themselves (in this case, the faces). Hearing that a person stole, lied, or cheated makes it more likely that we will actually see a structurally neutral but purportedly villainous face. While the magnitude of the effect might seem small in everyday terms (500 ms), it is a compelling demonstration of affect's completely top-down influence on vision.
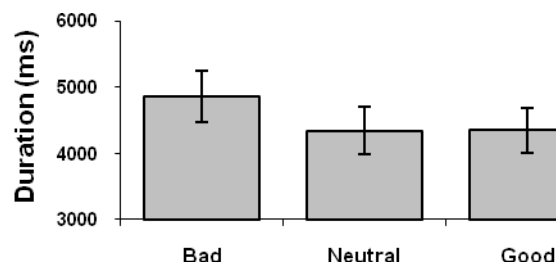
Figure 2. Duration faces were seen (Experiment 1).

In Experiments 2 and 3 (Anderson, Siegal, & Barrett, in preparation), we asked whether directly modifying the feelings of the perceiver influences what they see. To change participants' affective state, we presented evocative images that are known to evoke an unpleasant feeling (guns, fire), a pleasant feeling (puppies, flowers), or neutral feeling (buckets, rocks). Participants then saw faces (either angry, happy, or neutral) paired with neutral objects (houses) under binocular rivalry conditions. As predicted, we found that all faces dominated in participants who felt unpleasant; happy faces dominated in participants who felt pleasant. These findings indicate that what a person feels influences what he or she sees in a very fundamental way.

The Influence of Feelings on Blindsight

In Studies 4a-c (still underway), to understand the influence of feeling on detecting a fleeting threat cue, we had participants perform a signal detection task where on each trial they were presented with a more or less threatening face (ranging from angry all the way to neutral) very briefly (67 ms). Their task was to judge whether or not a face was threatening, and they earned points for correct judgments and lost points for errors; points were traded in for a reward at the end of the experiment. We first had participants perform this task in a baseline condition, and then repeated following an affect induction procedure where participants viewed pleasant, unpleasant, or neutral images while listening to music (Wen & Barrett, under review). Thus, judgments of threat were made under uncertainty ("more-" or "less-threatening" faces were difficult to tell apart from one another) and risk (costs and benefits were associated with incorrect and correct decisions). Under these judgment conditions, committing some number of mistakes is unavoidable; decision-makers must try to achieve an optimal blend of mistakes and correct categorizations in order to maximize the overall outcome. We examined how affect influenced this process.

Compared to the baseline condition, participants in a highly aroused unpleasant state performed better (earned more points during the threat detection task) than all other participants. This enhanced performance appears to have been mediated by an increase in sensitivity (when in a high arousal negative state, people were more sensitive to small changes in facial expression of the target faces, even though those faces were presented very briefly). Affective feelings did not influence bias (the tendency to judge a face as threatening or not), although people who performed better on the threat detection task used a more conservative threshold for judging when a threat was present (i.e., they said "yes" less often).

Feeling rarely changes in a way that is completely separate from environmental conditions, however. So next we examined how affective feeling influenced blindsight for threat when the external conditions for threat changes.

After again viewing images and listening to music to maintain their assigned affective state, some participants were next assigned to a situation where threats became less frequent (i.e., a change in base rate of threat) (Experiment 4a). Some were assigned to a situation where missing a threat became more costly than mistakenly judging a threat to be present that was not actually there (i.e., false alarms) (Experiment 4b). Some were assigned to a situation where the detection of threat became more ambiguous and therefore more difficult (Experiment 4c). In all cases, individuals who were feeling aroused but unpleasant calibrated best to the change (in terms of points earned), and this was mediated most consistently by changes in sensitivity: in a highly aroused but unpleasant affective state (typical of anxiety or anger), people were able to detect the presence of fleeting threat better, even under conditions of uncertainty and cost.

Contributions to Basic Science

First, this research demonstrates that a person's affective feelings from the body influence consciousness (for discussions, see Barrett & Bar, 2009; Barrett & Bliss-Moreau, 2009; Duncan & Barrett, 2007). The environment may literally look different to different people depending on how they feel at the moment.

Second, people vary widely in the intensity and variability of their affective feelings, so that some people may be affectively wired to see certain types of information better. It is also possible that new experiences can substantially change this wiring. Although individual differences are not the main focus of this proposal, the individual difference measures we collect will be used to indicate whether this is a fruitful avenue for further research.

Third, the current research also allows us to ask different questions about the accuracy of emotion perception (which is a core feature of emotional intelligence). Our findings suggest that a person's feelings cause them to be more perceptually sensitive to information that is contained on the face, but potentially at the expense of attending to information that is simultaneously available in the surrounding context. Context is crucial to perceiving emotion accurately and is routinely encoded during emotion perception (Barrett & Kensinger, in press). So it is possible that when people are feeling unpleasant and highly aroused, thus allowing a quick detection of threat, they may ironically be worse at the more fine grained analysis that is required for accurate emotion perception (in disambiguating fear and anger, for example). This hypothesis remains to be tested.

Finally, and perhaps more importantly, this research will inform an ongoing debate over the distinctiveness between emotion and cognition, and suggests that the distinction is not respected by the brain (Barrett, 2009; Duncan & Barrett, 2007). The most far-reaching implication of this work is that "thinking" (e.g., sensing and categorizing or deliberating on an object) might not be a fundamentally different sort of psychological activity than "feeling" (i.e., constructing a state to represent how the object affects you).

Potential Army/Military Applications

One challenge facing American Soldiers in contemporary combat and peace keeping situations is identifying whether another individual is a threat or not. Unfortunately, threats rarely announce themselves (and the same might also be said for positive information). The current research will help to examine when a person's momentary feeling state allows them to see threat better when (and only when) it is present.

The current research sets the stage for developing programs to train Soldiers to cultivate feeling states that can help them perform in varied combat situations. If the base-rate of threat changes in a given situation, Soldiers might be most effective at calibrating if they cultivate a highly aroused unpleasant state that will help them to see quick changes in the environment that predict danger (e.g., quick changes in other peoples facial muscle movements) or see threatening objects that easily blend into their background (e.g., hidden improvised explosive devices). We have yet to determine what sort of feeling state Soldiers might want to place themselves into so that they are less likely to perceive ambiguous visual information as being threatening. In some situations, for example, a Soldier might not want to mistake a cell phone for a gun.

Finally, being a good military leader means being able to read and influence the affective state of your troops. Individuals who are able to easily see "micro expressions" in their Soldiers or who can cultivate their affective state for optimal performance. As a result, the on-going research may offer a behavioral way of assessing whether certain individuals would make good leaders. Individuals who are particularly good at detecting micro expressions would be effective leaders in combat situations since they could use this information to estimate morale and guide their social interactions with other Soldiers appropriately.

Future Plans

To continue with experiments as outlined in our proposal.

References

Anderson, E., Siegal E. H., & Barrett, L. F. (under review). What you feel influences what you see: The role of affect in visual consciousness. Manuscript under review.

Anderson, E., Bliss-Moreau, E., & Barrett, L. F. (under review). The visual impact of affective information. Manuscript under review.

Barrett, L. F. (2009). The future of psychology: Connecting mind to brain. *Perspectives in Psychological Science, 4,* 326-339.

Barrett, L. F. & Bar, M. (2009). See it with feeling: Affective predictions in the human brain. *Royal Society Phil Trans B*, 364, 1325-1334.

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, 41, 167-218.

Barrett, L. F., & Kensinger, E. A. (2010). Context is routinely encoded during emotion perception. *Psychological Science,* 21, 595-599.

Duncan, S., & Barrett, L.F. (2007). Affect as a form of cognition: A neurobiological analysis. *Cognition and Emotion*, 21, 1184-1211.

Dissecting Situational Strength: Theoretical Analysis and Empirical Tests

Contract #: W91WAW-09-C-0096
Institution: George Mason University

Contract Dates: 8/20/2009 to 8/19/2010
PI: Reeshad S. Dalal
Co-PI/Subcontractor: Rustin D. Meyer
(Georgia Institute of Technology)

Problem(s)/Research Question(s) – Dispositional forces (e.g., personality traits) have been argued to predict behavior in "weak," but not "strong," situations. The current research presents a theoretically unified conceptualization of the construct known as "situational strength," provides comprehensive tests of its primary predictions, and assesses the effects of misfit between preferred and actual levels of situational strength on the job.

Technical Barrier(s) – Thus far, there has been little theoretical or empirical consensus regarding how to best conceptualize, operationalize, measure, and consequently empirically test the predictions associated with the psychological construct of "situational strength."

Significance/Impact for Basic Research – The present research provides, for the first time, a comprehensive test of the contention that the predictive power of dispositional forces (such as personality traits) vis-à-vis behavior varies as a function of the strength of the situation. In addition, the present research takes the situational strength literature in an entirely new direction by assessing employees' preferences for specific levels of situational strength, which in turn allows for an assessment of the deleterious consequences of a misfit between the preferred and actual strength of the situation.

Potential Transitions – Knowledge gained in this basic research may be usefully employed in:
- ARI's applied selection research programs (e.g., SELECT21)
- ARI's research on Selection, Classification, & Performance Metrics for the Future Force Soldier (STO IV.SP.2002.1)
- ARI's Occupational Analysis Office

Overview

Human behavior is ultimately a function of dispositional forces, situational forces, and their interaction. Although little consensus exists vis-à-vis the taxonomic structure of situations, leading the impact of situations on human behavior to be poorly understood, it has been argued (Snyder & Ickes, 1985) that the single most important property of situations is their "strength."

Situational strength is typically conceptualized as a general force that homogenizes enacted behaviors, thereby restricting the expression and predictive validity of dispositions. In "strong" situations, the most appropriate course of action is clear. In such situations, the effects of each person's unique dispositional profile are minimized. In "weak" situations, in contrast, the most appropriate course of action is not at all clear. In such situations, the effects of each person's unique dispositional profile are maximized. Stated differently, dispositional forces should predict behavior in weak situations but not in strong situations. For example, variability across persons in the level of conscientious behavior should be restricted in situations in which

conscientiousness is explicitly expected and rewarded (i.e., strong situations). There should, however, be much more variability across persons in conscientious behavior in situations in which conscientiousness is neither explicitly expected nor rewarded (i.e., weak situations). In the former case, conscientious behavior should be exhibited by virtually everyone, regardless of personality. In the latter case, conscientious behavior should be exhibited only by those with a conscientious personality.

The present research provides a comprehensive test of the contention that the predictive power of dispositional forces varies as a function of the strength of the situation. The present research additionally assesses employees' *preferences* for specific levels of situational strength, which in turn allows for an assessment of the deleterious consequences of a misfit between the actual and desired strength of the situation. Thus, the purpose of this research is to enhance our knowledge of situational strength by developing it into a more comprehensive theory of person-situation interactions, by more thoroughly testing its primary postulates, and by examining its effects from the perspective of person-environment fit. We believe that the idea of situational strength, if properly developed, represents the single most promising area of research into the impact of situations on human behavior.

Research Approach

The research approach involves employee surveys. As part of this research, we are validating a new survey measure of situational strength, then assessing whether situational strength influences the extent to which employees' dispositional tendencies predict employees' on-the-job behavior, then validating a new survey measure of employees' preferred levels of situational strength, and finally assessing the (presumably deleterious) impact of misfit between preferred and actual levels of situational strength. The primary data-analytic approaches used are factor analysis, convergent and discriminant validation, and moderated multiple regression.

Accomplishments

We have written items tapping the four facets of situational strength (derived in a previous paper; see Meyer, Dalal, & Hermida, in press, *Journal of Management*): clarity, consistency, constraints, and consequences. We have also collected data from over 100 research participants who performed a cognitive sorting task to assess the extent to which each of the situational strength items is associated with the intended facet as opposed to an unintended facet, multiple facets, or no facet. The next step, which will begin shortly, involves the collection of data that further assesses the dimensionality and functioning (specifically, what is referred to as "convergent and discriminant validity") of the new measure of situational strength.

Contributions to Basic Research

The research contributes to a better psychological understanding of the nature and structure of situations (a fundamental unresolved issue in psychology) as well as the manner in which situational forces alter the impact of dispositional forces on behavior. The expected products of our research are as follows: (1) a measure of situational strength and a measure of an employee's preferred level of situational strength, (2) conclusions regarding the extent to which situational

strength alters the impact of dispositional forces on behavior, and (3) conclusions regarding the extent to which misfit between actual and preferred levels of situational strength lead to deleterious outcomes.

The present research will also provide the foundation for several areas of future research.

One anticipated area of future research includes an investigation of short-term within-person changes in the momentary experience of situational strength. This could be accomplished via ecological momentary assessments of situational strength volatility over small time-units such as hours or days for the same person in the same job (or MOS). This line of research would allow researchers to better understand the influence of external forces on the moment-to-moment fluctuations in performance.

A second anticipated area of future research is investigating the impact of situational strength on constructs beyond personality. To take just one example, it is well known that job attitudes (e.g., job satisfaction, organizational commitment, and employee engagement) contain a substantial dispositional component. The relationship between job attitudes and job performance--which has been referred to as the "holy grail" of industrial/organizational psychology--has not yet been studied from a situational strength perspective, but we suggest (as does a quantitative review of the satisfaction-performance relationship; Judge, Thoresen, Bono, & Patton, 2001) that situational strength has the potential to increase our theoretical and empirical understanding of this relationship.

As a third, and potentially very influential, anticipated area of future research, numerous research projects could be generated by assessing situational strength in non-work settings. After all, personality is implicated in numerous types of non-work outcomes. To take just one non-work domain--namely, health--situational strength could, and should, be used as a "lens" to understand and increase individuals' adherence to medication schedules as well as behavior modification programs such as smoking cessation and weight loss. An example of an existing procedure, aimed at increasing adherence to medication schedules, that operates in accordance with situational strength principles is that women with a prescription for birth-control pills are asked to take one pill per day and are provided with blank pills on days when no pill would otherwise be taken.

Potential Army/Military Applications

The construct of situational strength has profound implications for the U.S. Army because, according to recent statements about the changing nature of the Army's role and structure (e.g., the 2007 Army Posture Statement and the 2006 Quadrennial Defense Review Report), Soldiers are now operating in environments that are more ambiguous, unpredictable and fluid than in the past, and Soldiers must now confront more potentially conflicting objectives than in the past. These new challenges are motivating an Army-wide (i.e., systems-level) emphasis on Soldier adaptability, flexibility, and uncertainty management. Stated differently, current and anticipated future Army situations are much "weaker" than those experienced traditionally—and the Army as a whole is re-orienting to reflect this decrease in situational strength.

The implications of situational strength on personnel selection and assignment in the Army are at least two-fold. First, the predictive power of dispositional forces should be higher today than when typical situations encountered by Soldiers were "stronger." This phenomenon highlights the continued need for the improved measurement and conceptualization of Soldiers' dispositions. In the ambiguous and uncertain situation facing the modern Soldier, individual discretion, decision-making ability, and affective stability will matter much more than in the past. Screening for dispositional characteristics such as Soldier personality is, therefore, likely to be more important as the Army continues to undergo changes that essentially weaken the situations that Soldiers typically experience. The Army is actively exploring the use of non-cognitive dispositional variables for Soldier selection and classification, and the current research not only serves as external confirmation for the usefulness of this endeavor but also provides a theoretical and empirical basis for more accurately estimating the predictive power of these dispositional variables.

Second, inherent to the idea of situational strength is the proposition that the predictive power of dispositional forces will not be uniform across MOSs, because the amount of situational strength that is typically experienced is likely to vary as a function of specific duties and circumstances. For example, the Special Operations Forces are much more likely to operate in weak situations wherein Soldiers are required to take direct action under conditions of extreme ambiguity, compared to MOSs such as the Quartermaster Corps, which are more indicative of strong situations because policies and procedures typically dictate appropriate behaviors. This has implications for Human Resource practices because it is unsafe to assume that the predictive power of dispositions will be uniform across MOSs. Despite the changing nature of the Army, differences across MOSs in terms of situational strength are likely to remain, and the current research offers an understanding of how and why these differences will affect the predictive power of dispositional forces. In fact, the ongoing occupational analysis efforts in the Army have the potential to benefit greatly through the incorporation of information regarding the situational strength associated with various MOSs.

In addition to focusing on actual levels of situational strength, we also focus on the levels preferred by individuals. A misfit in the amount of situational strength desired by incoming Soldiers, compared to the amount of situational strength that is typically present in the modern Army in general and/or specific MOSs in particular, is likely to lead to job dissatisfaction, psychological strain, and intent to quit (or to fail to reenlist). Of particular concern is the deleterious impact of the type of misfit involving Soldiers who desire greater situational strength than is provided by the situation. We therefore contend that Soldier retention and reenlistment are affected not only by phenomena such as stress on the battlefield and prolonged deployments, but also by the atypically "weak" situations engendered by the presence of flexible and adaptable non-state adversaries in modern conflicts. Thus, our current research examined leads directly to a line of applied research dedicated to mitigating the deleterious effects of situational strength misfit by classifying Soldiers into MOSs that are more appropriate to their needs.

Future Plans

As of mid-May 2010, we have collected and analyzed data that assesses the functioning (specifically, what is known as convergent and discriminant validity) of the scale measuring the level of situational strength associated with employees' jobs. We have also collected, and are currently analyzing, data assessing the extent to which situational strength alters the predictive power of dispositional forces vis-à-vis several dimensions of employee job performance. We have just begun to collect data assessing not only the functioning of the new scale measuring employees' preferences for situational strength, but also the extent to which a misfit between actual and preferred levels of situational strength is associated with deleterious consequences such as stress, dissatisfaction, and intent to quit. We expect to finish collecting data by the beginning of July.

Military Experiences in the Life Course

Contract #: W91WAW-09-C-0074   **Contract Dates:** 6/08/2009 to 6/07/2011
Institution: University of North Carolina, Chapel Hill   **PI:** Glen H. Elder, Jr.

Problem(s)/Research Question(s) – What are the consequences of recruitment pathways and attributes for military training, retention, and competence in the all-volunteer force? What life history factors make a difference in the post-service effects of military service?

Technical Barrier(s) – Lack of appropriate data sets on military service over the past ten years.

Significance/Impact for Basic Research – Understanding the life history outcomes of military service will allow for more targeted recruitment efforts and better support of military personnel following discharge.

Potential Transitions – Knowledge gained from this study could be used to:
- Enhance recruitment efforts for the all-volunteer force.
- Identify impact of a service member's life history to his/her military experience.
- Provide guidance for enhanced training for military personnel.
- Determine the work-life and health consequences of military experience.
- Determine areas of need in providing support to military personnel following their discharge from military service.

Overview

This research is designed to investigate the antecedents and consequences of military service, using linked data from the National Longitudinal Study of Adolescent Health (Add Health) and the Defense Manpower Data Center (DMDC). It will extend research begun with ARI contract W91WAW-07-C-0049 in two key ways. The project will take advantage of the 4th wave of Add Health data when study members are between 24 and 32 years old. This extension of the sample will enable us to investigate the behavioral and health outcomes of service in the all-volunteer force. With the return of overseas respondents, the original sample added a significant number of men and women with military service. The second benefit of the research is the ability to access critical information from the U.S. military archive (DMDC) which is not available in Add Health. Such information includes key events by age of service personnel, training activities, and overseas experiences.

Research Approach

By using data from Waves I, II, and III of Add Health, we have been able to identify distinguishing attributes and pathways associated with entry into the all-volunteer force. Nearly 1,000 respondents from Wave III reported military service (spanning all branches), either past or present. By comparing the sample in terms of two birth cohorts – a younger cohort of respondents in grades 7-9 in 1994, and an older cohort with respondents in grades 10-12 – we

can investigate the post-service health, adaptations, and accomplishments of men at different life stages.

Add Health was designed as a life course study but we have found it to have insufficient information on the timing of key events that bear upon military service in the lives of men and women. This has underscored the need to obtain military record data on the timing, training activities, and overseas military experience from DMDC. We view these records as essential for supplementing the military data in Add Health. They will enable us to address our research questions more thoroughly.

Accomplishments

As this contract began in June 2009, our accomplishments to date are limited and primarily relate to data definitions and use agreements. ARI's Human Use Committee gave approval to the project and its access to DMDC records on November 2, 2009. Release of final, cleaned Add Health Wave IV data has been delayed due to the extension of Wave IV preliminary fieldwork that was initially scheduled to end in September 2008 but was not completed until February 2009. The Wave IV tracing data file is currently being finalized and will then be sent to DMDC by the Add Health Data Security Manager for the data merge.

Contributions to Basic Research

This contract is in the beginning stages of the project and thus we cannot adequately address this question at this time.

Potential Army/Military Applications

- Inform policy on recruitment and its consequences.

Future Plans

- Wave IV picks up men and women who were overseas at the time of Wave III. Add these cases to the original sample, and repeat key analyses on pathways to the all-volunteer force.
- Make final revisions in papers for publication.
- Use Wave IV data to construct measures of military experience (e.g., exposure to combat) and post-service life patterns on work, education, family, and health.
- Initiate procedures for merging selected measures from DMDC with the Add Health military file.

Achievement in Complex Learning Environments as a Function of Information Processing Ability, Knowledge, and Self-Control

Contract #: W91WAW-09-C-0073   **Contract Dates:** 6/08/2009 to 6/07/2011
Institution: Case Western Reserve University   **PI:** Joseph F. Fagan

Problem(s)/Research Question(s) –The purpose of the present contract is to expand the pool of items used to test the validity of a culture-fair test of information processing ability and culture-fair measures of self-control for the prediction of academic and occupational achievement.

Technical Barrier(s) – Current measures of intelligence are, typically, estimates of knowledge that do not control for opportunity for past exposure to the information upon which that knowledge is based. As a result, current tests of intelligence have not proven to be culture-fair. Recent research on which the present proposal is based has shown that measures of information processing ability that experimentally control for past exposure have proven to be valid and culture-fair estimates of intelligence. The validity of an intelligence test based on information processing ability is based on predictions to achievement in academic settings and to success in the work place. Such achievement and success is determined by factors in addition to information processing ability, factors such as past knowledge, cultural background, motivation and attitudes. The extent to which each of these factors contributes independent variance to academic and occupational achievement remains to be determined.

Significance/Impact for Basic Research – This research will provide additional evidence for a theory defining intelligence as information processing ability and for a method to assess intellectual ability in a culturally fair way. We will explore the manner and extent to which motivational and attitudinal factors interact with intelligence and knowledge to effect performance in complex situations.

Potential Transitions – The results of the present research will *add fundamental knowledge to behavioral science... and... discover general principles* … [through an] *integrated programmatic effort to develop and to test theory* (as called for by RACO in BRO-BAA, page 1). The specific goal is to discover the relationships among information processing ability, knowledge, and self control that lead to achievement and productivity in complex situations. The goal is in keeping with the Army's aim to improve its ability to *Select, classify, train, and/or develop SOLDIERS and leaders who …are adaptable…*[and who]…*can function effectively in …information rich…environments* (BRO-BAA, page 1, II, 1, a, b). One of the BRO-BAA Basic Research Areas of Interest is *Leadership skills*. The research aids in the goal of identifying *Leadership skills in Adaptability…when faced with novel situations* (BRO-BAA, page 4, III, C, 1). A further area of interest that the research addresses is *Leader development.* In particular, *Maturity,* where the ARI BRO-BAA program calls for *developing a better understanding of the relationship between adult learning and growth in leadership ability.* The call is for *research to explore the applicability of …principles* [of learning and self-development] *to mature adults in their 30s and 40s* (BRO-BAA, page 5, III, C, 4).

Overview

What determines achievement in a complex learning environment? Is it how well new information is processed? Is it how much relevant knowledge is brought to the situation? Is it a sense of self-control? Guided by theory, the purpose of the proposed effort is to explore the manner and the extent to which learning ability, knowledge, and self-control interact to determine achievement. Specifically, college students and older adults 35 to 45 years of age, representative of the general population as to race and gender, will be tested for their ability to process new information, for the extent of their existing knowledge of word uses and meanings, and for their response to questions about their sense of self-control. These estimates of learning ability, knowledge, and self-control, respectively, will be employed in regression analyses along with race to determine whether and/or the extent to which each contributes independent variance to the prediction of scholastic achievement on the part of the younger sample, and to occupational status achieved by older adults.

Research Approach

Adults are given tests of knowledge of newly learned word meanings, sayings, similarities, and analogies as well as measures of recognition memory for novel faces and for abstract patterns to measure information processing ability. They also are given tests to measure their existing knowledge of word meanings, opposites, and analogies. All are asked to fill out a self-control survey and a questionnaire measuring their tendency to give socially desirable responses. A number of different self-control surveys will be given to see if there is a common factor regarding self-control. Regression analyses will assess the variance contributed by information processing ability, knowledge base, self-control (minus social desirability) and race to grades for 18-24 year old college students and occupational success for 35-45 year olds.

Accomplishments

Since the inception of the effort (6/8/09), some 300 college students and 198 older adults have taken part in the research. The results are in agreement with our past research. New learning of verbal information, past knowledge, and self-control are each predictive of class grades for the college students and of occupational success for the older adults. Moreover, both new learning of verbal information and self-control are, again, culture-fair with regard to the prediction of either academic or occupational success. Recognition memory for previously unknown cartoon faces adds variance independent of that provided by new learning of verbal information to the prediction of class grades for college students. The same recognition tests add no variance, independent of that provided by new learning of verbal information, to the prediction of occupational success for the older adults. Preliminary analyses indicate that recognition memory, unlike new learning of verbal information, may decline with age. Based on our progress so far, it is anticipated that testing, scoring, and analysis of the data from the total planned sample of 1,000 participants will be readily accomplished within the time limits of this program.

Contributions to Basic Science

The research will aid in further demonstrating that *intelligence is a function of an aptitude* [learning ability in novel situations] *that cuts across domains and how to develop a method for measuring this aptitude* [and to further] *validate* [this] *cognitive ability against relevant criteria*, (i.e., academic achievement and occupational success) (BRO-BAA, page 5, III, D, 1). The research will aid in understanding *personal motivational factors* [such as] *self-control...which are critical to success...* [by] *operating together* [with] *cognition… to achieve…in difficult situations* such as academic or occupational success, or, with regards to Soldiers, in *difficult military operations* (BRO-BAA, page 5, III, D, 2 and page 6, III, F).

Potential Army/Military Applications

Two of the guiding missions of the ARI both historically and currently are to provide new technology to meet the personnel challenges of the Army in recruiting, selecting and assigning quality people and to provide scientific insight into leader development. The present research provides basic scientific data to aid in *SOLDIER selection, assignment, and performance* and in *Leader Development.* In particular, findings from the current and proposed research may aid in the development of a *person-job match model* [which incorporates] *motivation as well as aptitudes.* Specifically, the results of the present research will aid the Army in measuring individual differences in cognitive ability and self control without regard to racial ethnic status. Using tests of information processing ability and self control known to be culture-fair may provide an increase in the number of eligible recruits and an incentive to re-enlistment on the part of minorities. The tests of new learning ability and the estimates of self control which emerge from the current research and from the proposed new research may be used in future research with adolescents and young adults interested in or involved in military service to add independent variance to that possible with standard tests now given by the armed services to the prediction of success in the military. Within the Army, an additional use could be the culture-fair prediction of achievement (e.g. rank) within a circumscribed age range and on success of new recruits in completion of basic training (and on drop-out rates).

Such tests may also be of use as a culture-fair addition to the current tests used for admission tests to the military academies. The use of such culture-fair tests may also prove valuable for hiring of personnel by the Army in various countries and for assessment of members of the military in various cultures, e.g. for selection of foreign military members for training in the USA. Tests of new learning ability can also be evaluated as to their efficacy for the early detection of, decline in, impairment of, or recovery of ability following injury, trauma, or illness. Such tests of new learning ability could be of use to Veterans Administration facilities and to Army hospitals (e.g., Walter Reed) for diagnosis and clinical evaluation of cognitive function, extent of neurological functioning, and to evaluate the effectiveness of rehabilitation techniques.

Specific transitions to applied research include:

Selection, Classification, & Performance Metrics for the Future Force Soldier (STO IV. SP. 2002.1). Findings from the present research on the role of information processing ability, knowledge, and self-control in achievement can be applied to a personnel selection and classification system geared [toward identifying Soldiers that are] self-directed and effective

information processors and problem solvers [a system that] might be incorporated into future enlisted selection and classification procedures [as well as being used as] outcome measures for future Soldier training and assessment.

Future Classification and Assignment. Findings from the research may aid in the development of a person-job match model [which incorporates] motivation as well as aptitudes.

Performance Measures for 21st Century Soldier Assessment. Findings from the research may be applied toward the development of a cost-effective…assessment system …for certifying a Soldier's qualification for job performance [and aid in the provision of] recommendations for an Army-wide system to certify the…qualifications of Soldiers for job performance.

Future-Train: Techniques and tools for C4ISR Training of Future Brigade Combat Team Commanders and Staffs. The research focuses on how to measure cognitive skills, knowledge and self-control which interact to influence performance in complex situations. Such aptitude and motivation may underlie the skills of future commanders and staffs.

Leader Development, Accelerating Leader Development. The proposed research may improve the assessment of cognitive skills, self-control, experience based learning, and the acquisition of tacit knowledge. Each of these factors may prove to be of importance to the selection and development of leaders.

   The implications of the proposed research for society in general are similar to those for the Army. Selection of applicants for admission to higher education is a prime example of how research sponsored by the ARI can solve societal problems. Currently a sizable gap in SAT or ACT scores exists between students of majority or minority racial ethnic status. Attempts at affirmative action by institutions of higher learning to increase minority admissions have been successfully challenged in legal cases brought before state and federal courts. One solution to the disparity in admission rates to college or to employment due to race or economic status is to develop culture-fair, valid tests of academic and occupational aptitude that can be seen as race-neutral alternatives. The current research supported by ARI adds to the evidence that brief, reliable tests of information processing based on the ability to acquire new information as well as estimates of self control are valid predictors of academic and occupational achievement and do not discriminate between groups differing in racial ethnic or in economic status. As was the case with the development of the Army Alpha, such a current program of research on race-neutral alternatives puts the ARI, once again, at the forefront of the solution to important societal problems.

Future Plans

During the remaining 18 months of the contract period, a further goal is to produce tests of new learning ability that, in addition to being culture-fair, are training proof. This will be done by giving new learning ability tests for sayings, vocabulary, similarities and analogies and recognition memory to young adults at two points in time to show that performance does not improve by practice on alternate forms. This will be done in two ways. The first is by creating alternate forms of new learning ability tests for sayings, vocabulary, similarities and analogies as well as for recognition memory that are given to young adults at two points in time with an interval between tests of about two months. An alternate form of a previously given test will be created in which the same words to be learned are used but which are now, on the second occasion (two months later) endowed with meanings previously associated with other words in

78

that set. The same tactic will be used for tests of recognition memory. In both cases (learning word meanings or recognizing faces), previous experience would not help but might hinder performance on the second test.   To provide a baseline measure of difficulty level, a control group of young adults would be tested only once, but on the second test given to the repeatedly tested participants. Practically, finding that tests of new learning ability are training proof may aid in alleviating the disparity in admission rates to colleges based on scores on standard tests of knowledge, scores which can be increased by training programs available to those of higher economic status. Currently, the culture-fair, predictive tests of recognition memory noted above are unfamiliar cartoon faces. A second goal is to expand the pool of recognition memory items using unfamiliar abstract patterns. The aim, again, is to obtain additional, culture-fair items for the prediction of achievement; items that can be easily employed in future research in any cultural context. A third aim is to go beyond the use of the one measure of self-control used in the current research and to explore the predictive validity of a variety of measures of self-concept. To do so, young adult college students will complete five scales of self-control, self-efficacy, and self-confidence. Factor analyses will be employed to obtain the common factor(s) accounting for their responses and correlations will be obtained between factor scores and academic grades. The goal is to find items with regard to attitudes toward the self that are most predictive of achievement. A final focus as to test standardization will be pilot research with both college students and older adults (workers within universities or hospitals will constitute the possible older adult population of interest) aimed at the creation of computer based tests of new learning ability and estimates of self-control that can be taken via a web site, scored on line, and that can provide immediate feedback on performance. All of these goals can be accomplished within the time remaining.

Cognitive Task Complexity and General Intelligence

Contract #: W91WAW-09-P-0209   **Contract Dates:** 6/15/2009 to 6/14/2010
Institution: Otterbein College   **PI:** Meredith Frey

Problem(s)/Research Question(s) –Is it possible to develop nonverbal cognitive tasks that predict outcomes as well as more traditional, verbally-loaded tasks? Also, will these tasks demonstrate lower racial/ethnic group differences than more traditional tasks?

Technical Barrier(s) – Traditional measures of intelligence, which demonstrate good predictive validity, incorporate learned information that is susceptible to environmental influences. The basic cognitive elements that underlie intelligence-- and consequently performance on those traditional intelligence tests-- have been difficult to measure and are not well understood.

Significance/Impact for Basic Research – Research into a new battery of cognitive tasks has far-reaching applications. In addition to increased understanding of the mechanisms of intelligence, these tasks can allow for assessment of cognitive capacities that is independent of acculturated knowledge.

Potential Transitions– Knowledge gained in this basic research may be usefully employed in:
*   Selection, Classification, & Performance Metrics for the Future Force Soldier.

Overview

For over one hundred years, researchers in the field of human intelligence have been attempting to understand the structure of ability. Some argued that intelligence is one thing that impacts performance in a variety of areas, while others advocated for a more complex structure of intelligence (see Jensen, 1998 for a thorough review). One theory suggested that intelligence is actually a system of separate but related processes, some of which are central (necessary for all tasks), others peripheral (Detterman, 1987). If this were correct, then a fuller understanding of individual and group differences in intelligence would be possible only if these separate processes could be identified. Furthermore, no matter what theory or model of intelligence one supports, it is only useful if it predicts important human behavioral outcomes, such as job performance or academic achievement.

Earlier efforts to identify some of the central processes in intelligence yielded mixed results. If a single task could be changed to have a higher correlation with general ability, then that change could be said to be accessing an additional central process in the intelligence system. Several attempts to do just that have failed. Adding complexity to basic reaction time and pattern matching tasks resulted in increases in the time it takes to solve a problem, but not in the correlation that task has with general ability (Frey & Detterman, 2002; Frey & Detterman, 2004). One experiment, however, showed that when individuals tried to find a pattern that *almost* matched a sample, the amount of difference and the type of difference between the two patterns produced changes in both time to solve and correlations with general ability (Frey, 2007).

The ultimate goal of the above research was to develop a battery of cognitive tests that measure more accurately the central processes of intelligence, without requiring any verbal learning or acculturated knowledge. The elementary cognitive tasks proposed in this research

did not rely on knowledge of vocabulary, formal mathematics training, or other topics that could be influenced by environmental factors, such as school quality. It was expected that these tests would demonstrate lower racial/ethnic group differences while maintaining the ability to predict important outcomes. However, there were problems with the previous research that made firm conclusions regarding the structure of abilities difficult to obtain. First, the process of sampling university students, particularly at the private research university where the above experiments were carried out, did not allow for an examination of the tests in minority populations. Second, it was not possible to know how participants were working to solve the individual problems. In cases where a pattern almost matched a sample, people could have been operating on either the sample pattern or on the pattern that was supposed to be the match. Finally, participants in the previous research always received tasks in the same order where the patterns on the first set of tasks were more discrepant than they were on the second set of tasks. Because of this order, it was possible that the earlier tasks took longer to do because participants were getting used to the procedures, but that these earlier tasks were not actually good measures of participants' abilities.

To address these limitations, a new experiment was proposed that recruited participants from a more diverse population, presented tasks in random order, and asked participants to explain how they worked to solve individual tasks. Specific hypotheses follow:

H1. A randomized presentation of near-match trials will demonstrate the effect of establishing context on individual differences in pattern recognition (that is to say, tasks presented second will demonstrate higher correlations with general ability, regardless of degree of discrepancy).

H2. Consistent with Luo and Petrill (1999), the cognitive task battery will predict academic achievement in a college sample.

H3. Group differences in latent constructs measured by the cognitive task battery will be negligible or significantly lower than those associated with traditional intelligence tests.

Research Approach

All participants are recruited from colleges and high schools in central Ohio. Participants represent both sexes, a broad range of general ability, are primarily between the ages of 17 and 25, and contain students who self-identified as white (Caucasian) and as members of various minority groups. All students complete a 36-item nonverbal test of general reasoning ability, the Advanced Progressive Matrices (Raven, 1962). After completing the reasoning test, participants begin a computer-based task of decision-making ability. Participants are instructed to look at a sample pattern and select— from five alternatives— the pattern that matches the sample (see Figure 1, below). In the first set of trials, the match is exact; however, in the last four sets of trials, the sample does not have an exact match. Rather, participants are instructed to find the pattern that almost matches the sample. Finally, participants are asked to sign a release allowing the experimenter to access their academic records, specifically their grade point averages (GPA).

Accomplishments

Data collection is still ongoing. So far, 94 students, 31 males and 63 females, completed the computer-based and written portions of the experiment. Unfortunately, some important information, such as academic achievement data, is not available for roughly half of the participants who have not yet completed the first quarter of their freshman year in college. The data that are available, however, partially support the above hypotheses. Consistent with

previous work by Frey (2007), tasks where the participants are instructed to find a pattern that almost matches a sample take longer to complete, and the amount of time taken is effected by how different the sample pattern is from the near match. So, if a participant is supposed to find a match between two patterns that are different by two squares (Figure 1 c & e), that takes longer than finding a match between two patterns that are different by only one square (Figure 1 b & d). Other findings are summarized by hypothesis, below.

a. Standard Match-to-Sample   **b.** Match-to-Sample Plus 1   **c.** Match-to-Sample Plus 2

d. Match-to-Sample Minus 1   **e.** Match-to-Sample Minus 2
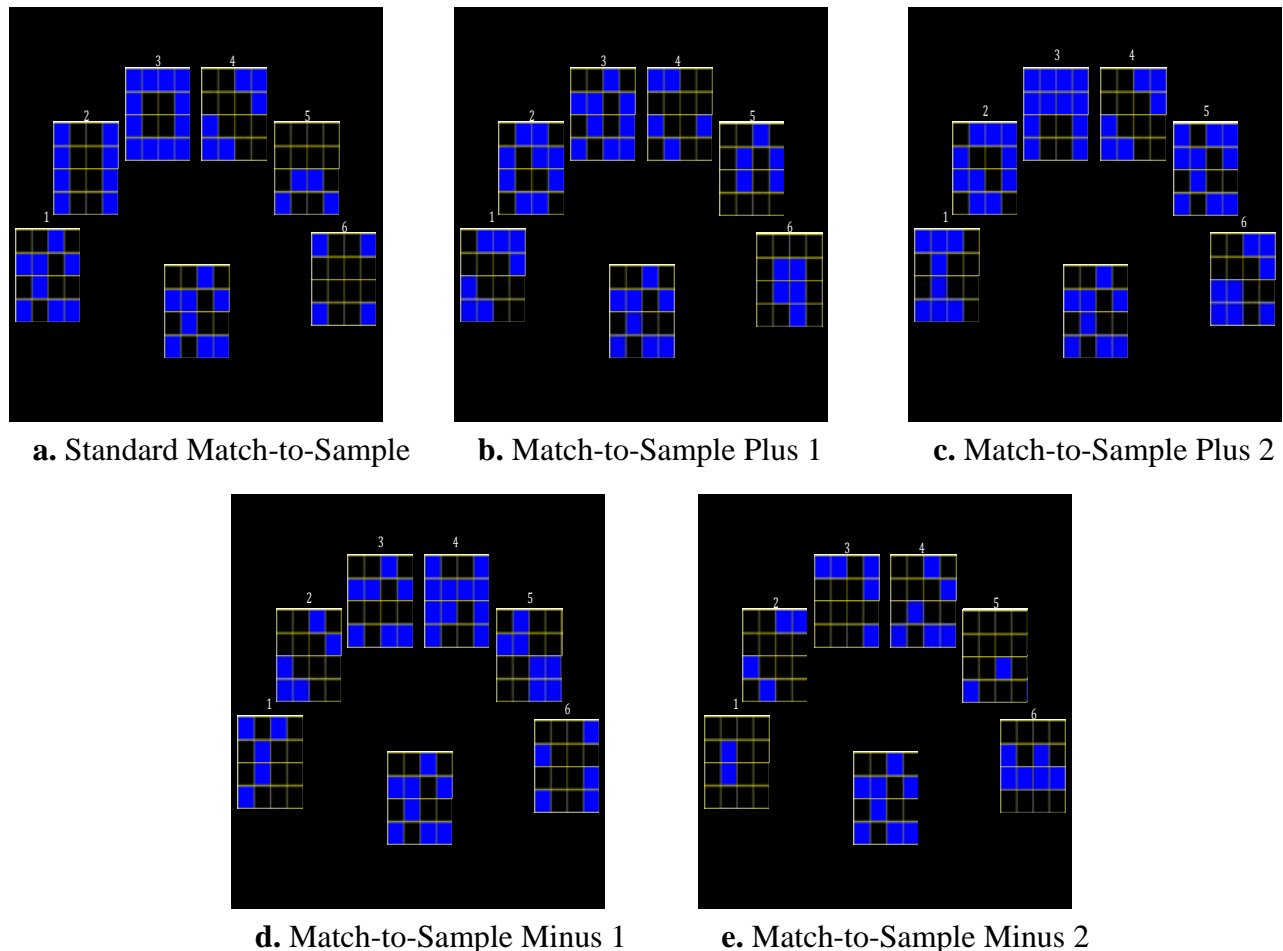
Figure 1. The match-to-sample tasks. In the standard condition (**a**), an exact match exists between the probe (bottom center) and target (position 1). Plus 1 and Plus 2 tasks (**b** and **c**, respectively) require the participant to mentally add 1 and 2 squares, respectively, to the probe to make it match the target. Minus 1 and Minus 2 tasks (**d** and **e**, respectively) require the participant to mentally subtract 1 and 2 squares, respectively, from the probe to make it match the target.

       H1: When examining the effect of the order participants received, it is difficult to draw any firm conclusions due to a small sample (there were 24 possible orders, and many participants had data excluded for high error rates). If participants receive more discrepant trials first, they are roughly 400 to 500 milliseconds faster on the less discrepant trials than participants who

receive the less discrepant trials first. In these cases—where participants received more discrepant trials first— the correlation between speed and general ability was also higher for the later trials than for the earlier trials. However none of these correlations is significantly different from zero, and the fact that they are positive—meaning that faster performance is related to lower scores on general ability—is unusual (see Figure 2). Therefore, it does not appear that learning the rules that govern the task, and using those rules to solve later trials, is a likely candidate as a central process in intelligence. Receiving less discrepant trials first does not systematically affect performance on later trials, but it does result in statistically significant correlations with general ability that are more consistent with previous research.
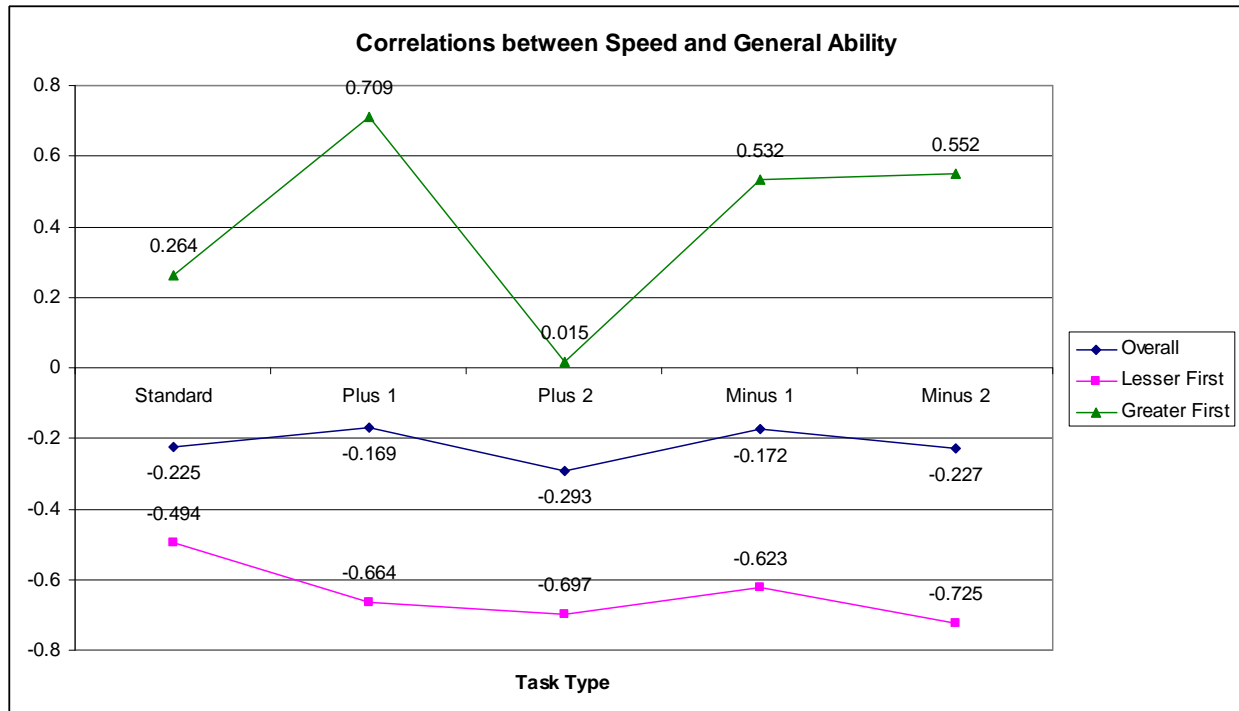
Figure 2. Mean decision time x general ability correlations, separated by order of tasks. It is clear that the participants who received tasks with lesser discrepancy (Minus 1 and Plus 1) first demonstrated speed x general ability correlations that were consistent with previous research (moderate in size and negative in direction). In contrast, those participants who received greater discrepancy trials first (Minus 2 and Plus 2) had non-significant, but positive speed x general ability correlations. The reason for the difference in direction is unknown.

H2: Again, because of roughly half of the sample does not yet have academic achievement data, it is difficult to interpret the results of correlations between speed on cognitive tasks and GPA. However, some preliminary results are quite interesting. It appears as though receiving less discrepant trials first results in significantly higher correlations with GPA. These correlations indicate that, in this very small (N=5) sample, speed on some of the tasks almost perfectly predicts GPA (see Figure 3).

Correlations between Speed and GPA

Figure 3. Correlations between decision time and college GPA, separated by order of tasks. It is again clear in the above figure that subjects' performance on tasks predicts outcomes (in this case academic achievement) well, but only if participants receive lesser discrepancy trials (Plus 1 and Minus 1) first. However, as GPA data was not available for roughly half of the participants, and error rates on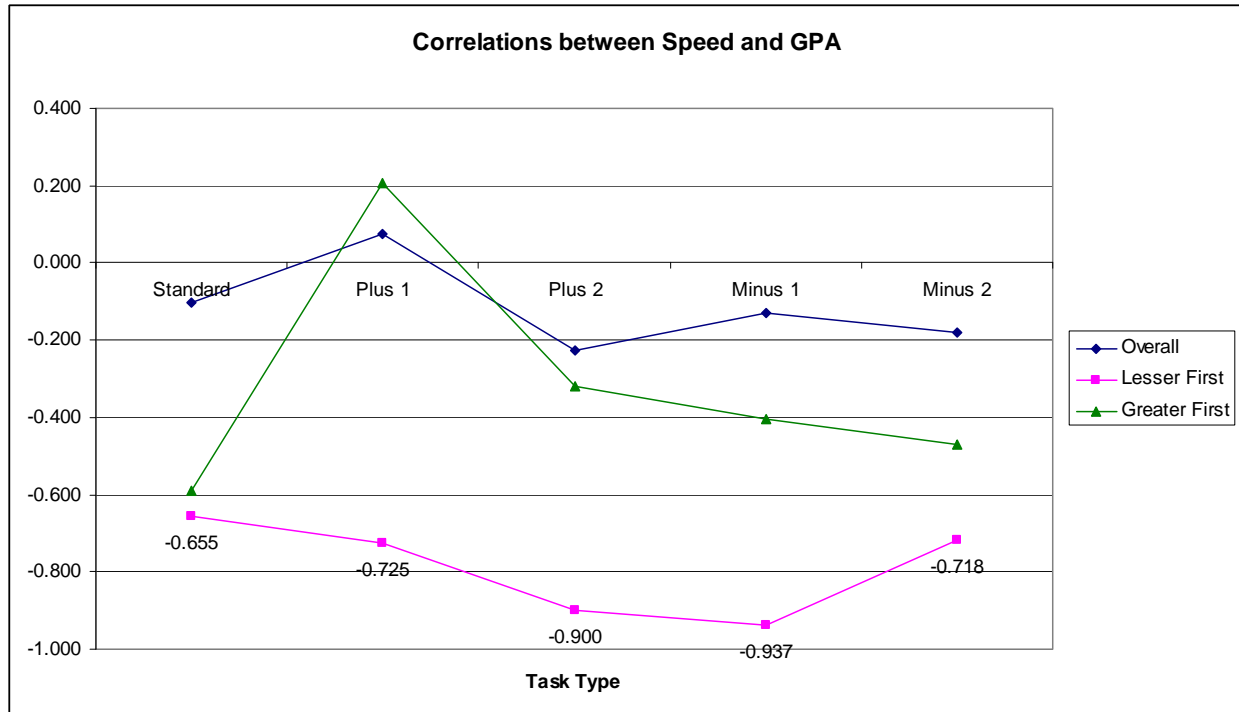 some tasks were quite high, sample sizes in some of the conditions above were very low, and any interpretation must be regarded as preliminary.

H3: Although more data are required to perform an analysis of latent constructs, some information is available that indicates that performance on these basic tasks is not subject to the same racial/ethnic group differences as other more traditional tasks. In the data collected so far, participants who self-identified as belonging to one or more minority groups had significantly lower scores on the test of general ability (Raven's Advanced Progressive Matrices) and on their college entrance exams (composite scores on the ACT). However, these same participants had slightly higher GPAs than their non-minority classmates. Performance on the basic computer tasks indicates a slight speed advantage for white students, but this difference, consistent with the hypothesis, is smaller than the size of the difference in more traditional measures.

Contributions to Basic Research

The data collected thus far are a step towards developing a battery of nonverbal cognitive tasks that have the potential to predict important outcomes. The primary criterion that intelligence tests have attempted to predict has been academic achievement, and, at least under some circumstances, the tasks developed here predict that achievement as well as more traditional tests. These findings indicate that it is possible to measure potential without assessing acculturated knowledge.

Potential Army/Military Applications

According to the Broad Agency Announcement (W91W-08-BAA-RACO), the U.S. Army seeks unbiased assessment of cognitive abilities. The Armed Services Vocational Aptitude Battery, although a good measure of *g,* demonstrates racial group differences on the mean scores as well as on the prediction of meaningful criteria in a large representative sample (Tiu, 2001). Although these data are from an older sample, before the removal of numerical operations and addition of assembling objects subtests, many of the subtests of the ASVAB that rely heavily on verbal information remain unchanged.

The elementary cognitive tasks developed in this research do not rely on knowledge of vocabulary, formal mathematics training, or other topics that can be influenced by environmental factors, such as school quality. However, it remains important that the battery demonstrate predictive utility. This research attempts to objectively evaluate models of cognitive ability across groups against academic achievement, the primary criterion used to establish the predictive validity of traditional psychometric intelligence tests.

It is imperative that the U.S. Army be able to quickly identify those individuals who possess the cognitive skills to succeed in the military environment. Assessing Soldiers and recruits on the basis of central processes allows for an objective evaluation of potential without incorporating environmental factors. Because the battery is more culture-fair, it is likely that a new, computer adaptive battery of cognitive tasks will continue to show much smaller, even insignificant racial and ethnic group differences. If group differences do become evident, the more targeted tasks allow researchers and program developers to directly measure the cognitive processes responsible for the differences and, ideally, develop training methods to remedy such differences.

Future Plans

Data collection, particularly of academic achievement data, is ongoing. Once these data become available, the results presented above will be reanalyzed with increased statistical power. Another important potential predictor is strategy. Participants were asked to indicate how they worked to solve trials where there was no exact match to the sample pattern. These data have yet to be examined in light of their effect on performance. Finally, the pattern-matching task described herein is not the only task to be included in the eventual battery. Work has begun on other task types, particularly auditory tasks and tasks of working memory, that can be combined with the pattern matching task to develop a more complete assessment of basic abilities.

Facial Affect Reciprocity in Dyadic Interactions

Contract #: W91WAW-08-C-0024
Institution: San Francisco State University

Contract Dates: 03/01/2008 to 02/28/2010
PI: David R. Matsumoto

Problem(s)/Research Question(s) –
- How does the synchrony in facial expressions of emotion exchanged between group members while engaged in cooperative and competitive interactions relate to decision making and team performance?
- Is the relationship between facial synchrony and team performance moderated by time constraints or cultural differences between interactants?

Technical Barrier(s) –
- The lack of a theoretical framework for understanding the interpersonal functions of facial expressions of emotion across time.
- Little empirical research exists on how team performance and decision making changes as a function of human emotion and its interchange between interactants.

Significance/Impact for Basic Research –
- The results of this study will contribute to understanding the social meaning of facial expressions of emotion and the role they play in group performance tasks and decision making.
- These results will contribute to the further development of a theoretical framework that gives a central role to the exchange of facial expressions of emotion in group performance tasks.

Potential Transitions – Knowledge gained in this basic research may be usefully employed in:
- The development of tools designed to aid individuals in recognizing emotions in others in social interactions, with the potential for use in group decision making tasks.
- The development of tools designed to aid individuals in understanding how to use emotional displays in social interaction in an adaptive way to achieve goal-desired outcomes, and help leaders and team members in maximizing team performance.
- Systemic changes in the nature of communication technologies available to operations-level decision making. If such facial signs can be detected, leaders in such situations can be trained to recognize such signs and intervene to maximize team performance (assuming the incorporation of teleconferencing in team-based decision making operations).
- The development of emotion–based lexicons for use in netcentric environments that can aid in online interactions beyond emoticons.
- Provide crucial information concerning individual-level adjustment and adaptation in multicultural teams that can be used for the development of leader training programs and tools.

Overview

The purpose of this research is to investigate a new concept, Facial Affect Reciprocity (FAR) and its role in contributing to interaction and performance in groups. FAR is proposed as a new theoretical framework to guide empirical studies about the interpersonal functions of emotion. More specifically, we hypothesize that FAR is an important mediating mechanism that influences group performance and function. FAR refers to the *exchange* of facial expressions among group members across time, and the *linkage* between specific types of emotion. It is based on the research demonstrating that facial expressions of emotion have enormous signal value and thus considerable interpersonal function, i.e., they coordinate social interactions. FAR also proposes that specific *sequential combinations of emotions exchanged among group members* signal important information about the nature of the interaction and relationship. Nine different sequential combinations of emotion (e.g., cooperation, betrayal, competition, etc.) that can occur in a dyadic interaction are described and hypotheses that these predict team performance better than data from individual facial expressions of emotion are proposed.

Research Approach

In order to observe the facial expressions of emotions group members exchange while involved in task performance and decision making, pairs of research participants play a modified version of the Prisoner's Dilemma game. Prisoner's Dilemma is a game with a long history of research in psychology and behavioral economics that typically elicits strong emotions in the players. Players are presented a set of choices that they must make independently and simultaneously that result in gains or losses of payoffs for themselves and their partner. Three study conditions are being conducted to observe possible differences in FAR based on the nature of the relationship between the players and the conditions of play. Participants may be assigned to (1) a partner who is a stranger and from the same cultural background or (2) a stranger and from a different cultural background or (3) the same condition as (1), but the participants must play the game under severe time constraints.

The research is being conducted at a major urban university with a culturally diverse student population. The general research approach is an experiment in a laboratory setting. Prior to playing the game, participants complete a set of personality and emotion expressivity measures. Participants' faces are videotaped in order to record the facial expressions of emotions exchanged between the players while they play the game. These videotapes are scored for the number and type of sequential combinations of emotions that are exchanged between the pair and finally analyzed in relation to the individual and group game performance outcomes.

Accomplishments

Data collection has been completed for the main study, Study 1 (N = 120). All self-report and performance data have been cleaned. Preliminary analyses have indicated that the degree of synchrony between the two players' self-reported emotional states are correlated with a number of performance indices for both the dyad and the individuals (Table 1). Notably, the more synchronous their emotions, the greater the actual amount of cooperative plays both engaged, the less competitive plays, and the greater the payoffs for both of them.

Table 1 – *Correlations between Synchrony in Self-Reported Emotions and Dyad Performance*

Outcome	Pre-test	Post-test
Dyad Collaboration	-.137	.295*
Dyad Ruination	-.036	-.505***
Dyad Impasse	.078	-.364**
Dyad Destruction	-.047	-.327*
Cooperative Plays, Player A	-.028	.455***
Competitive Plays, Player A	.127	-.355**
Cooperative Plays, Player A	-.023	.356**
Competitive Plays, Player A	.106	-.240*
Total Payoff Player A	-.034	.324*
Total Payoff Player A	-.040	.442***
Total Dyad Payoff	-.038	.394**

We are currently in the process of coding the facial behaviors of the participants, using the Facial Action Coding System (FACS). This is a very time-intensive, laborious process; coding for the 11 minute videos from the experiment typically require approximately 10-15 hours of work. To date we have completed the coding for 41 participants, and have conducted some preliminary analyses on a subset of these data. We computed a score to represent the degree of synchrony between the two participants' facial expressions of emotion, and then correlated this facial synchrony measure with various dyad-level performance measures. As you can see below (Table 2), preliminary analyses indicate that the synchrony between the two participants' facial expressions of emotion does signal something meaningful and important about the behavioral outcomes of their task.

Table 2 – *Correlations between Synchrony in Facial Expressions of Emotion and Dyad Performance*

Outcome	Correlation
Dyad Collaboration	.900**
Dyad Ruination	-.830*
Dyad Disconnection	-.769*
Total Payoff Player A	.838*
Total Payoff Player A	.716*
Total Dyad Payoff	.785*

We have begun the data collection for Studies 2 and 3 and have collected data from approximately 20 dyads in each study. Data collection has been hampered by restrictions placed on the psychology participant pool this semester because of mandatory furloughs and cuts to available sections that have occurred because of budget problems faced by the State of California and passed onto San Francisco State University.

Contributions to Basic Science

This research will contribute to the current literature on the interpersonal function of emotion. Specifically it will extend knowledge about the signal value of facial expressions of emotion from simple one-way sender and receiver modalities to the more complex interactions between sender and receiver across time. In addition, the research may provide new knowledge about how cultural diversity and stress influence interpersonal functioning in groups.

Potential Army/Military Applications

Effective team performance is vital for all branches of the military at all levels. An understanding of how FAR influences team functioning and performance, and how cultural diversity and stress within teams may influence FAR, provides the Army/Military with data to make personnel selection decisions, group assignments and improve training and assessment materials. The data will likely suggest the importance of being able to see participants in cooperative task environments, either in face-to-face interactions or with the use of technology when appropriate (e.g., videoteleconferencing). The findings may suggest the development of sensor technologies that can detect when teamwork is breaking down, as well as the development of training tools to help individuals monitor and recognize the signs of team cooperation or breakdown.

Future Plans

- Data collection for Studies 2 and 3 need to be completed.
- Facial coding for Study 1 needs to be completed, and then Studies 2 and 3 when those data are finalized.
- Final data analyses need to be performed when facial coding for each study is completed.
- Final reports need to be drafted.

The Development and Validation of a Situational Judgment Test to Predict Attrition Incrementally Over General Cognitive Ability and a Forced-Choice Personality Test

Contract #: W91WAW-07-C-0014   **Contract Dates:** 03/26/2007 to 03/25/2010
Institution: Work Skills First, Inc.   **PI:** Michael A. McDaniel

Problem(s)/Research Question(s) – Will a situational judgment test predict attrition-related criteria over and above general cognitive ability and a forced choice measure of personality? Can the situational judgment test be made resistant to faking?

Technical Barrier(s) – Lack of an objective scoring key for situational judgment items, feasibility of measuring personality with a situational judgment test using knowledge instructions, faking of self-reported information in high-stakes testing, difficulty in building scales given that situational judgment items typically measure more than one trait at a time.

Significance/Impact for Basic Research – Understanding the extent to which maximal performance assessment of personality and judgment, through a situational judgment test with knowledge instructions, can be used to improve better prediction of attrition-related variables.

Potential Transitions – Knowledge gained in this basic research may be used in:
- ARI's applied selection research programs (e.g., SELECT21).

Overview

There are several ways to address attrition. One method is to increase incentives for troops to stay beyond their eligible retirement by providing higher rates of pay with greater years of service. Another method is to provide reenlistment bonuses to those considering leaving the military. An additional method for reducing attrition is to develop selection methods designed to predict retention. This last method is the topic of the proposed research.

This project entailed the development of a situational judgment test designed to predict attrition in the U.S. Army. We framed our proposal in the context of providing incremental prediction over and above the ASVAB and AIM for the criterion of first-term attrition. However, a measurement instrument that can do this would likely yield some prediction for other criteria such as reenlistment, productivity, adaptability, and performance in semi-autonomous environments. More importantly, the method proposed offers a blueprint for the development and scoring of construct-oriented situational judgment tests to target such criteria.

Although applicants for the military seldom have substantial job experience, we argue that military applicants have knowledge areas that are gained in one's initial job and through life experience gained in family, peer or school interactions. We suggest that measures of these socialization knowledge areas might show more variance among applicants for entry-level positions than among individuals with substantial experience. McDaniel, Whetzel, Hartman, Nguyen and Grubb (2006) offered a model of situational judgment and job performance indicating that situational judgment tests likely tap general job knowledge. By this, they refer to

91

basic knowledge common to most jobs and might be viewed as work socialization knowledge. These knowledge areas would include the value of showing up to work, getting along with co-workers, dressing appropriately, following the directions of your supervisor, and refraining from employee theft and inappropriate language. We assert that these knowledge areas can be readily assessed through situational judgment tests.

However, users of any self-report measure, including situational judgment tests, need to consider the effects of faking when tests are taken in high-stakes situations, such as screening into the military. Faking can be viewed as a distinction between typical and maximal performance. Cognitive ability tests, job knowledge tests, and work samples are maximal performance tests. Respondents know that they are being evaluated in a high-stakes situation that requires them to perform at their best. Thus, when applicants for the military take the ASVAB, they know that they should use their knowledge to select the best answer.

Most personality tests have typical performance instructions. Consider the following instructions:

> Please use the rating scale below to describe how accurately each statement describes *you*. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. (Drawn from http://ipip.ori.org/newIPIPinstructions.htm.)

Thus, personality tests solicit a respondent's typical behavioral tendencies. However, applicant screening situations are high-stakes, maximal performance situations. Thus, some applicants may choose to ignore the behavioral tendency instructions and respond instead with this maximal performance instruction:

> Use your knowledge to identify the best answer, that is, the answer that will most likely get you the job.

Thus, when applicants seek to fake their answers to employment tests, they are using knowledge instructions rather than behavioral tendency instructions.

Situational judgment tests are unusual in that response instructions are sometimes knowledge instructions (maximal performance instructions) and sometimes behavioral tendency instructions (typical performance instructions). Below, we review these types of instructions.

Knowledge instructions ask respondents for the correct answer. Knowledge instructions would include:
- Pick the best response
- Pick the best response and the second best response
- Pick the best response and the worst response
- Rate each response on its effectiveness.

Behavioral tendency instructions ask respondents for their typical responses. Behavioral tendency instructions would include:

- What would you most likely do?
- What would you likely do and what is the second most likely response you would do?
- What would you most likely do and what would you least likely do?
- Rate each response on the likelihood that you would do the response.

McDaniel and Nguyen (2001) relied on a metaphor to argue that knowledge instructions would be faking resistant and behavioral tendency instructions would be readily faked. They argued that situational judgment tests with behavioral tendency instructions are similar to personality tests because both instruments ask for a respondent's typical behavior. Situational judgment tests with knowledge instructions are similar to knowledge tests because both types of tests ask for the correct or best behavior.

Consider the personality item: "I keep my work area neat and clean." This item is not unlike a situational judgment item related to how a person might respond if they find that their work area is disorganized and cluttered. For both the personality item and the situational judgment item, a disorganized but honest respondent is likely to report that they are typically disorganized. However, a disorganized respondent engaging in faking would report that he or she is organized.

Consider this psychometric knowledge item: "What is the best reliability estimate to use with a construct heterogeneous measure?" A knowledge situational judgment item might ask the respondent to judge the effectiveness of various responses to a problem at work. For both the psychometric knowledge item and the situational judgment item, the respondent is being asked to rely on their knowledge for the correct or best answer. Both honest respondents and those seeking to fake would be motivated to provide the same answer, that is, the correct or best answer. A respondent lacking knowledge of the correct response, might guess but they cannot fake. Thus, we assert that situational judgment tests with knowledge instructions are faking resistant.

The findings of Nguyen, Biderman and McDaniel (2005) support our assertion. They evaluated the faking resistance of the same situational judgment test using both knowledge and behavioral tendency instructions. Although respondents could improve their scores with behavioral tendency instructions, the same test items with knowledge instructions were resistant to faking.

Research Approach

Situational judgment items were written to reflect school and work experiences that would likely be familiar to military applicants. Items were written to sample a broad range of experiences. A group of graduate students rated the item responses to the extent that they tapped the Big 5 personality traits and content related to quitting behaviors (e.g., giving up, not finishing what one starts). These ratings were used to identify a set of responses that might be readily categorized into Big 5 scales. These responses and their associated stems were assembled into a situational judgment test with 20 scenarios and 136 item responses. Other instruments in the

survey package included a demographics and background survey, a cognitive ability test, and a forced-choiced personality test. Attrition-related criteria included a biodata measure of quitting behavior and whether the respondent had received a high school diploma. For those with military service, we solicited information on whether the respondent completed a tour of service. After deleting random responders and surveys with large amounts of missing data, we had data on 702 respondents.

The analysis of the year-one data showed that the situational judgment test predicts attrition related criteria. We identified a subset of items with predictive utility and are collecting year-three data with the shortened test. This administration includes a financial incentive to perform well. We seek to deterimine if the situational judgment items retain their validity when respondents are motivated to fake to improve their scores.

Accomplishments

Our analyses to date involved:
- Understanding the structure of the biodata quitting measure.
- Examining the criterion-related validity of various scoring methods for the situational judgment test.

Each of these three topics is addressed below.

Understanding the structure of the biodata quitting measure

The research to date on understanding the structure of the biodata quitting measure is summarized in the following conference paper:

Fluckinger, C., Snell, A.F., & McDaniel, M.A. (2009, April). *Construct-oriented development of a biodata scale of quitting behaviors*. Paper presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology. New Orleans.

After the conference, the paper was submitted to the *International Journal of Selection and Assessment*. The paper received a "revise and resubmit" editorial decision and is being revised for resubmission to that journal.

Based on a review of the literature relevant to quitting, items were written to assess hypothesized dimensions of quitting behavior including past quitting, quitting as a stress response, and engagement/embeddedness. Quitting as a stress response is related to hardiness, or also commonly referred to as resilience. Quitting in response to stress is also associated with perceived organizational support (POS), which involves perceptions that the organizations people are involved with value their contributions and care about their well-being. Job embeddedness is the overall level at which one has become entrenched in a job. Embeddedness is not only caused by job-related factors, but also social factors, such as links to the community and family ties.

Based on exploratory and confirmatory factor analyses, we found that quitting behaviors and antecedents of quitting behavior were classified into five primary dimensions: past quitting behavior, perseverance, interdependence, commitment, and coping. This structure, in addition to the finding that the quitting dimensions are not directly caused by a higher-order latent quitting factor, leads to the conclusion that there are a range of social and work-related behaviors that may be valuable in addition to previous quitting behavior in predicting future quitting behavior. In short, the act of quitting or disengaging from a given situation is a complex combination of factors such that push or pull individuals.

Examining the criterion-related validity of various scoring methods for the situational judgment test

Our research to date on understanding the criterion-related validity of the various scoring methods for situational judgment test is summarized in the following conference paper submission:

McDaniel, M.A., Psotka, J. & Legree, P.J. (2009, April). *Toward an understanding of situational judgment item validity*. Paper presented at the 24th Annual Conference of the Society for Industrial and Organizational Psychology. New Orleans.

Scoring situational judgment test items has much in common with profile matching such that the mean item ratings (collected from experts or examinees) define the scoring profile, which is used to assess individual profiles (i.e., item ratings from individual respondents). A respondent's score on the test is a function of the degree of match between the respondent's answers and the group means. Cronbach and Gleser (1953) conceptualized profile matching with respect to elevation, scatter, and shape. Elevation is the mean of the items for a given respondent. Scatter reflects the magnitude of a respondent's score deviations from the respondent's own mean. Cronbach and Gleser argued that one should consider whether elevation and scatter are important in their profile matching application.

We examined three scoring methods. One method (raw consensus scoring) did not control for elevation and scatter. One method (standardized consensus scoring) controlled entirely for elevation and scatter. The third method (dichotomized consensus scoring) largely controlled for elevation and scatter. Our analyses supported two conclusions that have implications for improving situational judgment test validity. First, scoring methods that controlled for elevation and scatter yielded higher response option validities than methods that did not. The raw consensus scale validity was .06, compared to .33 for the standardized consensus scoring, and .34 for the dichotomized consensus scoring. Second, as shown in Figures 1 to 3, there was a U-shaped relationship between response option means and item criterion-related validity, such that items with low and high mean ratings were more valid than items with means near the mid-point of the Likert scale. This permits one to drop many of the items with middle-range means in order to substantially reduce test length without harming validity. For the biodata quitting criterion, the standardized consensus scale and the dichotomized consensus scale showed incremental validity over and above cognitive ability and a forced choice personality measure ($R = .56$ to $.58$ versus $.55$).
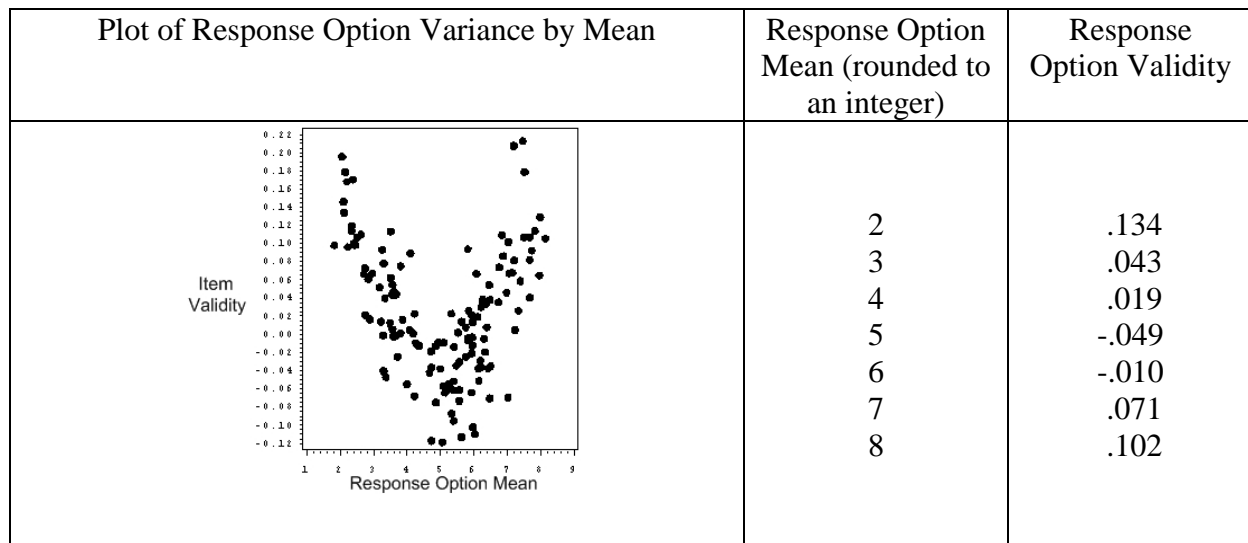
Plot of Response Option Variance by Mean	Response Option Mean (rounded to an integer)	Response Option Validity
	2	.134
	3	.043
	4	.019
	5	-.049
	6	-.010
	7	.071
	8	.102

Figure 1. Relationship between response option mean and response option validity using raw consensus scoring.

Plot of Response Option Variance by Mean	Response Option Mean (rounded to an integer)	Response Option Validity
	2	.167
	3	.122
	4	.133
	5	.054
	6	.074
	7	.142
	8	.149
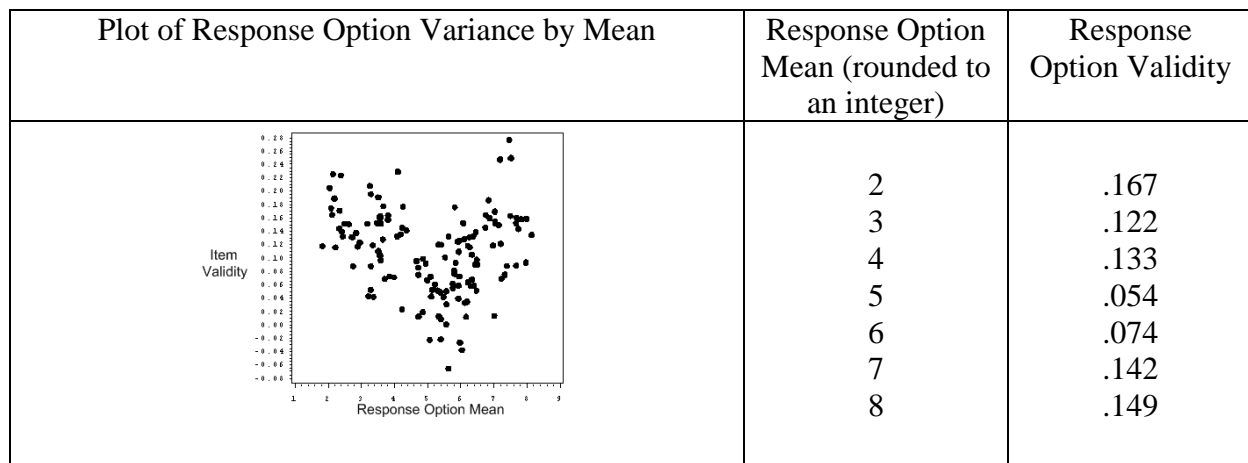
Figure 2. Relationship between response option mean and response option validity using standardized consensus scoring.

We also demonstrated that the scoring methods that controlled for elevation and scatter, also reduced mean White –Black score differences. Thus, for scoring SJTs using Likert scale response formats, one can simultaneously increase validity and reduce White – Black mean differences. We replicated this finding with a private-sector sample using a different situational judgment test. A paper demonstrating the effects in the Army sample and replicated in the private-sector sample was submitted to the *Journal of Applied Psychology*. The paper received a "revise and resubmit" editorial decision and is being revised for resubmission to that journal.

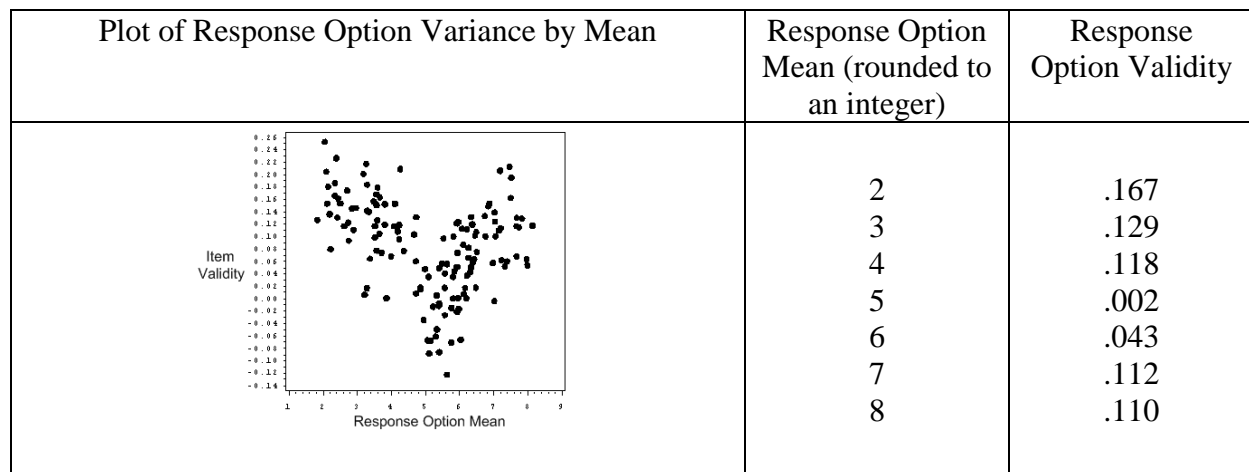| Plot of Response Option Variance by Mean	Response Option Mean (rounded to an integer)	Response Option Validity
	2	.167
	3	.129
	4	.118
	5	.002
	6	.043
	7	.112
	8	.110

Figure 3. Relationship between response option mean and response option validity using dichotomized consensus scoring.

Contributions to Basic Science

To date, this project has developed basic research knowledge in three areas. First, the project has identified five primary dimensions of quitting/attrition behvaior. These are past quitting behavior, perseverance, interdependence, commitment, and coping. The project has also developed a scale to assess these dimensions of quitting behavior. Second, this project has substantially advanced our knowledge of the factors that control situational judgment item validity. Specifically, we have demonstrated that it is important to control for elevation and scatter when scoring situational judgment items. This control substantially improves validity. In addition, we have identified a U-shaped relationship between item means and item validity, such that items with either high or low means have substantially higher validity. We have evidence that situational judgment tests can predict quitting behavior over and above cognitive ability and a forced-choice personality measure. Third, we have demonstrated that scoring methods that control for elevation and scatter simultaneously reduce White - Black mean differnces in test score performance.

Potential Army/Military Applications

Our research shows that attrition can be conceptually related to five distinct dimensions: past quitting behavior, perseverance, interdependence, commitment, and coping. These dimensions can be used to focus personnel screening efforts and to guide military job redesign. Our research has also shown the situational judgment tests can predict quitting behavior and incrementally predict over and above cognitive ability and a forced-choice personality measure. The validity is enhanced when one controls for elevation and scatter. Screening time can be substantially reduced, without harming validity, by dropping items with mid-range mean ratings. Also, because of the reduction in White - Black mean differences on the SJT occurs in conjunction with improved prediction, more Black military applicants can be classoified as suitable for service.

Future Plans

   The result of our research to date supports quitting behavior as a multi-dimensional construct. With the second wave of data collection, we can evaluate the extent to which situational judgment items are faking resistant. A demonstration that one can improve prediction accuracy with a faking resistant test and simultaneously reduce White – Black mean test difference will be very rewarding.

Multimedia Assessment of Emotional Abilities: Development and Validation

Contract #: W91WAW-07-C-0025   **Contract Dates:** 05/14/2007 to 05/13/2010
Institution: Educational Testing Service   **PIs:** Richard D. Roberts and Ralf Schulze

Problem(s)/Research Question(s): Can limitations associated with existing measures of emotional abilities (EA) be overcome by developing multi-media assessments and using innovative methodologies and advanced statistical models? How are these new measures related to meaningful outcomes such as coping with stress, decision-making, teamwork, and leadership?

Technical Barrier(s): Many current measures of EA use self-report techniques, which overlap with existing personality measures and are easily faked. Performance-based tests of EA are more promising, but existing measures are limited. These measures have problematic scoring keys, lack ecological validity, and have questionable psychometric properties.

Significance/Impact for Basic Research:

- The proposed research will result in the development of a set of methods for measuring EA. These methods could be used for assessing EA, either for personnel selection and classification or for assessing progress in EA training. In addition, a taxonomic model of EA and other affective constructs becomes possible.

- Understanding the role of EA in relation to other psychological constructs will provide a more comprehensive theory of human capability and performance.

- Cognitive abilities have proven valuable in predicting a range of outcomes valued by society; this research will document the incremental predictive validity afforded by EA and will identify outcomes that are particularly related to emotional abilities.

Potential Transitions: Knowledge gained by this research and development effort may be useful for:

- The Army Research Institute's (ARI's) applied selection research programs (e.g., Selection, Classification, & Performance Metrics for the Future Force Soldier; Performance Measures for 21st Century Soldier Assessment);

- ARI's work on cross-cultural competence. Identifying and understanding emotions in different cultural contexts appears an important skill for soldiers deployed to foreign theaters.

- ARI's applied training research programs (e.g., Future Force Warrior Training). Multimedia assessments may be retooled for this purpose and could, in principle, be a type of formative assessment;

- EA has been claimed to represent a set of constructs that are pivotal to successful leadership. Thus, another possible application is for ARI's leader development programs (e.g., Developing Leaders in A Changing Army, Assessing and Developing Leaders for Special Operations Units).

Overview:

Are some people better at managing their emotions and using this ability to make better decisions as well as interpret and understand the emotions of others? These are the essential

questions of research into emotional abilities (EA, also referred to as emotional intelligence or EI). This project is concerned with how we can determine whether some people are better at using their own emotions and understanding others. The project is also concerned with developing methods that reliably and validly measure differences in the ability to interpret, use, understand, and manage emotions, both in one's self and in others.

The importance of this research becomes clear in light of the purported outcomes associated with elevated EA (see Mayer, Roberts, & Barsade, 2008; Roberts, Schulze, & MacCann, 2008; Zeidner, Matthews, & Roberts, 2008). Individuals with high EA are thought to make better decisions under conditions of stress, to be better team players, to make better leaders, to refrain from damaging health behaviors, and to have more fulfilling relationships; in short, to be more adaptive to their environment. The current research seeks to establish a reliable method for measuring EA and to provide validity evidence for the concept. Part of this demonstration involves relating EA measures to established indices of emotion-related behaviors, intelligence, and personality, and in line with theoretically predicted relationships (i.e., positive relationships with emotion and intelligence measures, weak relationships with personality, etc.). Another part of this demonstration involves relating EA to meaningful outcomes (e.g., coping with stress, leadership). Acknowledging limitations in extant measures of EA, a feature of this research is the development of new, multimedia (video- and text-based) test procedures to assess these emotional abilities. In addition, each measure may be classified according to a taxonomic model that we have developed in recent research (Mayer et al., 2008).

Research Approach:

Tests of EA were constructed and given to samples of community college and university students (726 participants in total), along with measures of personality, ability, and outcomes. The methodologies used to develop the EA tests included: (1) A situational judgment test (SJT) paradigm (where participants rated a scenario for emotional relevance and/or salience); (2) an empathic agent paradigm (EAP, where event-emotion contingencies in others had to be perceived and recognized and emotion-behavior contingencies inferred from observed behavior to predict future behavior); (3) various information processing paradigms (with emotions as stimuli); and (4) an implicit association technique (where an individual's implicit association of emotions with words and situations was assessed).

All of the tests were computerized, where (1) and (2) comprised video sequences as constituent components, (3) and (4) included both figural and verbal elements. The use of multiple media was essential for several reasons. First, it enabled the assessment of certain EAs that cannot easily be assessed with other methodologies that are limited by the delivery medium. Second, it broadened the conceptualization of EA beyond measurement methods that draw excessively on verbal ability, a problem that has plagued the field to date. Lastly, it served to enhance the degree to which the test paradigms mirror the real environment. To give the reader a feel for the set of tasks described in (1) and (2) above, a screen capture from one SJT item is shown in Figure 1a. Figure 1b shows how the measure requires users to rank-order the quality of potential responses to an emotionally laden situation. This rank-ordering process requires a direct comparison of choices, and allows users to review the responses until they are comfortable with the quality of each response.

We evaluated the psychometric properties of these new measures, including tests for measurement invariance (i.e., is the test measuring the same things in different sub-samples) and

examination of sub-group differences (e.g., ethnic groups). Estimation of structural relations between latent variables underlying both the new and established (EA) measures served to test a proposed three-factor taxonomic model and provide evidence for the new measures' convergent and discriminant validity. Further validity evidence was collected by estimating the new tests' relationships with outcome measures (e.g., GPA, coping with stress, social support) where intelligence and personality variables were controlled.
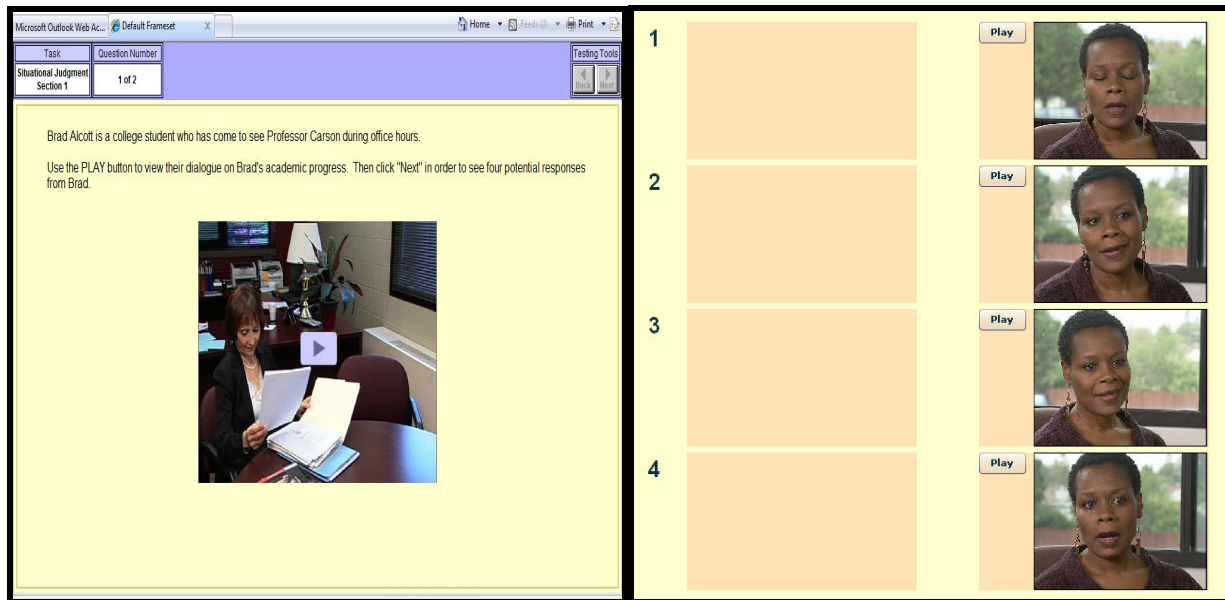

Figure 1a (left panel): Interface for the multimedia SJT. Figure 1b (right panel): The drag and drop interface used in the multimedia SJT (which was also used for the EAP measures).

Because the data were supportive of the new measures' quality we examined a broader criterion space, focusing initially on psychological well-being (i.e., how EAs relate to positive affect, as assessed by the Day Reconstruction Method--a process that uses recall of events throughout a day to remember emotions attached to those events [Kahneman et al, 2004]). Further, we examined additional psychometric properties of the multimedia EA instruments – specifically test-retest reliability – by giving these to a subset of participants who completed these tasks on the first occasion.

Accomplishments:

The data attest to the reliability, validity, and thus ultimately the usefulness of these new multimedia assessments of EA. In the initial study of 726 students, both the SJT and EAP measures were found to have acceptable internal consistency reliability (Cronbach alphas of .70 and .80, respectively). In addition, both of these new tests exhibited moderate to high relationships with another measure of EA – the Mayer-Salovey-Caruso Emotional Intelligence Test (Mayer et al., 2008) – with average correlations ranging from 0.25 (Perceiving Emotions) to 0.52 (Understanding Emotion). EAP and SJT also correlated $r = 0.50$ with each other, and $r = .31$ with a veridically scored measure of Emotion (Face) Perception, the Diagnostic Analysis of Nonverbal Accuracy Scale -2 (Zeidner et al., 2008). The SJT and EAP also had moderate to high correlations with disparate measures of cognitive ability; the highest correlation being between

EAP and a measure of crystallized ability ($r = .58$), and the lowest between the SJT and a measure of quantitative ability ($r = .27$). Neither measure was too highly correlated with personality, though both the EAP and SJT shared moderate correlations with Agreeableness (average $r = .30$). Finally, regression analyses suggested that cognitive ability, Conscientiousness, and the multimedia assessments consistently added to the prediction of grade point average (GPA). Collectively, these results mirror some of the major findings from recent meta-analyses of maximum performance measures of EA (e.g., Roberts et al., 2008), and suggest that the new measures appear to be assessing a construct that encompasses both emotional and intellectual components, and provide validity evidence for these EA test scores.

As noted previously, we followed up 131 of these participants to obtain test-retest data, in the process expanding the range of outcomes that these measures might predict. The test-retest reliabilities were .74 for the EAP and .54 for the SJT, values experts consider highly satisfactory. In addition, both the SJT and EAP correlated moderately (i.e., around $r = .30$) with indicators of psychological well-being, with the EAP also again predicting grades in the new semester ($r = .25$). Of great interest, perhaps, is the relationship that both the SJT and EAP share with indicators of net positive affect taken over a period of time using the Day Reconstruction Method (DRM). These correlations exceeded $r = .40$ for reported net positive affect while socializing and working and were noticeably lower for reported affect while in class, and lowest when, for example, spending time on the internet or eating food (though always still positive in sign).

These findings indicate we have seemingly been successful in developing multimedia assessments that are face valid and have acceptable measurement properties. Relationships with personality, intelligence, other emotions measures, and valued outcomes are all as might be expected in the light of theory (see Zeidner et al., 2008) and indicate that the scores have demonstrable construct validity (Roberts, MacCann, Matthews, & Zeidner, in press). The DRM results are highly suggestive for the practitioner; this approach is used in economic models and indicates that EAs may play a role in worker (and by extension, Soldier) satisfaction and productivity, for example. Clearly, there is a need to continue collecting data with these new assessments, but thus far they have shown considerable promise.

Contributions to Basic Science:

The EA measures under research and development use a variety of different methodologies: situational judgment, implicit association, information processing, reasoning, and empathic agent paradigm approaches. Current research indicates that each of these methodologies offers a new method of assessing EA, in a more objective and valid way. This research will also provide insight into the domain of EA, potentially providing an empirical taxonomy for placing tasks, allowing more research in the area, and informing our understanding of how humans use emotions in making decisions. Linking these EA assessments to other domains, such as teamwork and leadership also holds promise of more compelling models of human performance. Additionally, being able to differentiate between those who are able to appropriately recognize, manage, and use their emotions (especially in highly charged emotional situations) will enable organizations to recognize which individuals are likely to perform well in emotion-laden situations.

Potential Army/Military Applications:

The research we are conducting can be used in the ARI's applied selection and assessment research programs, with an eventual segue into use for classification, selection,

placement, and promotion of personnel. Additionally, information regarding EA may be useful in training related assessments for existing personnel. In particular, certain specialties and occupations may find that using EA as an assessment tool for the purposes mentioned above can increase Soldier success. As an example, ARI may find that Psychological Operations, Adjutant General's Corps, and Military Intelligence personnel with higher EA perform more efficiently and to a higher standard than those with lower EA. The relationship between higher EA and performance of duties required by these branches is not known at this time, but for personnel needing to make evaluations of intent and gauging emotional or social situations, such as might be encountered by these personnel, it is likely that higher EA will provide substantial benefits. Additionally, personnel in command positions may benefit from higher EA. Our research enables us to first develop a measure of EA and then validate that measure, so that its relationship with valuable outcomes is known.

As a training tool, a measure of EA could be used to identify personnel that need increased understanding of how to work with emotional situations, and then provide training to those individuals so they are able to better work with, understand, and manage their own and others' emotions. The developed SJTs and EAPs also might serve as training tools in their own right, with the potential for these to serve a formative assessment function. A presumed advantage of developing these EA assessments relies on their purported freedom from biases seen in traditional cognitive measures (i.e., they hold promise of correcting adverse impact). In addition, these measures might provide potential predictive information regarding a range of behaviors, which cognitive measures do not cover. These features may make these attractive noncognitive supplements to future versions of the ASVAB as recommended by Drasgow et al. (2006).

Future Plans:

We are currently completing a number of manuscripts related to the findings from the first two studies, and have also submitted papers to national and international conferences. Plans are also underway to conduct a study that will use a tactical decision-making task to investigate whether multimedia measures of EA predict better decision-making in negative emotional states. We predict that more emotionally able individuals will make better decisions under conditions of stress than their less emotionally able counterparts.

References:

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical Review of the Armed Services Vocational Aptitude Battery (ASVAB).* Technical Report FR 06-25. Alexandria, VA: HumRRO.

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The Day Reconstruction Method (DRM). *Scienc*e, *306,* 1776-1780.

Mayer, J., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507-536.

Roberts, R. D., MacCann, C., Matthews, G., & Zeidner, M. (in press). Emotional intelligence: Towards a consensus of models, measures, and applications. Social & Personality Psychology Compass.

Roberts, R. D., Schulze, R., & MacCann, C. (2008). The measurement of emotional intelligence: A decade of progress? In G. Boyle, G. Matthews, & D. Saklofske (Eds.), *The Sage handbook of personality theory and assessment* (pp. 461-482). New York: Sage.

Zeidner, M., Matthews, G., & Roberts, R. D. (2008). *What we know about emotional intelligence: How it affects learning, work, relationships, and our mental health*. Cambridge, MA: MIT Press.

Social Trends and Social Change in the United States: Impacts on Army Manpower, Personnel, and Operations

Contract #: W91WAW-09-C-0077   **Contract Dates:** 9/28/2009 to 9/28/2013
Institution: University of Maryland   **PI:** David R. Segal

Problem(s)/Research Question(s) – What impact does ongoing social change in the United States have on the Army?

Technical Barrier(s) – Social change is an ongoing process, and the ways in which social institutions and members of the population adapt to it, are processes that basic researchers in the behavioral and social sciences must constantly monitor in order to understand the contemporary world. The trends to be analyzed in this project represent changes in the social context within which the Army operates that differ from the past.

Significance/Impact for Basic Research – Understanding how organizations in general, and the Army in particular, respond to social change, will lead to more adaptive models of organizational phenomena such as leadership, professionalism, and cohesion.

Potential Transitions Knowledge gained in this basic research may usefully feed into ARI research programs in areas such as officer accession, professionalism, and leadership.

Overview This research will examine a range of types of data on social trends and social changes in the global system, American society, and the Army, including changes in the demography of the American population and labor force, interpreted in the context of sociological theories on systems, organizations, professions, motivations to serve, and the work/family interface. This research should inform processes of Army socialization, recruitment, retention, and performance.

Research Approach We will undertake analysis of social trends and changes These include a range of survey data on military personnel, ranging from surveys of the Army that date back to World War II (available through the electronic data collections of the National Archives, located on the University of Maryland campus), and including tailored surveys of specific Army units conducted by scientists at the Center for Research on Military Organization in recent years, as well as recent surveys conducted of military personnel and their families by the Defense Manpower Data Center. We do not envisage new data collection efforts in the Army as part of this effort.

Surveys to measure trends in the civilian sector include an ongoing collaboration with scientists at the University of Michigan to analyze the military-relevant data in the Monitoring the Future (MtF) program. Using these data, we have been conducting research on changing youth attitudes as they apply to the military since 1975. MtF is a cohort-sequential survey that collects data from about 16,000 high school seniors each year, and follows up sub-samples for 14 years after high school graduation. This has enabled us to analyze the correlates of propensity to serve, and to follow the trajectories of high school graduates as they choose between military

service, civilian employment, or higher education. Thus it has served as a basis for comparing young men and women who serve with their peers who do not. These analyses will be continued.

We have also developed a template for tracking age cohorts over time using Public Use Microdata samples (PUMS) of the decennial U.S. Census. In our current research we are following the World War II generation through the census data from 1950, 1960, 1970, 1980, 1990, and 2000. The template allows us to identify the characteristics of who serves and who doesn't, compare socio-demographic groups, and identify sequels to service, including educational levels attained, occupational status, and economic well-being, and to compare those who served with those who did not over the life course. In this project, we plan to apply the template to those who have served in the volunteer force. At the current time this would cover the first three decades of the volunteer force

These major quantitative data sources will be supplemented by other ongoing survey efforts as appropriate. These include the General Social Survey, conducted annually by the National Opinion Research Center with National Science Foundation support, the National Longitudinal Survey, and a number of survey programs tapping educational attainment and economic well-being of the population that will enable us to analyze military personnel in the context of the general labor force. Many of these data sets are available at the University of Maryland. Our location in the Washington area eases access to other federal data source.

Our location also gives us easy access to archival data sources. Some have been turned over to us. For example, we have a copy of the social network data on the Baathist organization accumulated by the 4th Infantry Division prior to the capture of Saddam Hussein, and continue to analyze these data to gain a better understanding of the social structure of an insurrection. In addition, the University of Maryland Library is a Federal Records Repository, and gives us access to government publications, with which our library staff is familiar. More generally, our Library has an excellent collection of social science serials, allowing us to track the most recent research on organizational change and evaluate its relevance to understanding military organization and the military profession.

Accomplishments We are only two months into this research effort. However, we are well into analyses of how professions have been changing in America, and how this impacts specifically on the Army as a profession. We have also undertaken an analysis of how America's response to the attacks of 9/11/2001 have affected Army manpower and personnel. Professional research publications on these two efforts will be submitted to ARI during the next six months.

Contributions to Basic Research This project contributes to our understanding of how changes in global and societal systems affect specific organizations, professions, families, and individuals that operate within them. By looking at the Army as a social system that contains within it all of the processes that occur in the larger host society, it serves as a laboratory for understanding dynamic social processes.

Potential Army/Military Applications No organization or profession operates in a social vacuum. All are acted on, and respond, to their environments. Understanding the dynamic interchange that goes on between the Army and its host society will assist Army personnel managers in understanding and managing the motivations and behaviors of their personnel, the stresses that relocations and deployments place upon Soldiers and their families, and the dynamic

processes that take place in military units (and in all work groups) such as cohesion and leadership that affect individual and unit performance.

Final Summary

We will continue to monitor and analyze changes in the demographic composition of the American population and of the Army, with particular reference to minority representation in both. We will continue to study how the social roles of women are changing in modern industrial nations, and how this affects the roles they play in military organization. We will continue to focus on the interface between work and the family in these societies, and the management of the demands placed on military families that function at this interface. And we will continue to track data on military personnel, both active and reserve, to better understand the role that military service plays in the life course of those who serve, and how it affects their transition to other social roles.

BRU RESEARCH OBJECTIVE #4: PROVIDE FUNDAMENTAL KNOWLEDGE FOR UNDERSTANDING AND EVALUATING ORGANIZATIONAL BEHAVIOR AND NETWORK SCIENCE RESEARCH

Research under this objective is focused on using social networks, whether in simulations, games, or Army organizations to help understand and predict group behavioral processes. Modeling and understanding the networks underlying human knowledge systems is also part of this objective.

Improved Data Extraction and Assessment for Dynamic Network Analysis

Contract #: W91WAW-07-C-0063   **Contract Dates:** 08/07/2007 to 08/31/2012
Institution: Carnegie Mellon University   **PI:** Kathleen M. Carley

Annual Report 2008-2009

Problem(s)/Research Question(s) – The Army has a need for rapid and accurate analysis of unformatted data in order to support various operations. Examples of raw texts include open source, human intelligence (humint), and situation reports such as those entered in Command Post of the Future (CPOF). For example, there is a need to identify texts and extract from them information about various actors and associated political, economic, military, social, and infrastructure related activities. Key questions include: How can text data be rapidly processed to extract meaning and not just key entities and content analysis?; How can links be extracted?; and How can the attributes of extracted entities be identified?; How can we infer sentiment?

Technical Barrier(s) – Dynamic network analysis tools can be used to process unformatted data. However, these tools need to be improved so that the time spent by the user in locating, ingesting, and coding data is minimized and so that analyses can be quickly upgraded as new information is collected. Entity extraction, particularly as it applies to names and locations, can be done quite well. Link extraction and inference is now possible, but is prone to errors if only proximity is used. Barriers to rapid development here include the lack of adequate tagged data for system testing, lack of appropriate unsupervised learning algorithms, and lack of common ontology for sentiments.

Significance/Impact for Basic Research – We propose to augment the DyNet tool suite so that a more flexible, automated, user-friendly, rapid, and accurate set of techniques are available for data extraction, coding, and assessment with particular focus on link extraction, attribute extraction, semi-automated ontology creation, and sentiment extraction. We focus on raw texts. Our initial focus was on link extraction, anaphora resolution, and semi-automated thesauri construction. These tools have been developed. They support the DOD need to assess socio-cultural and political information, rapidly identify key influences and opinion leaders. We focused this year on improving data processing speed, improving and refining link extraction, enabling attribute extraction, and sentiment extraction. These technologies support rapid assessment of cultural geography for the DOD. In addition, they enable assessment of early indications of change in sentiment by populations at risk and key opinion leaders. AutoMap serves three key purposes: identification of hot topics, assessment of trends in attitudes, and extraction of meta-networks to support DIME/PMESII modeling and social network analysis.

Potential and Actual Transitions – AutoMap, which we are extending under this project, is currently an underlying system in VIBES, which is being developed for the G2 and is now part of the HSCB program. This has been reviewed for incorporation in MAP HT and MAP HT extensions are being built. AutoMap can also be used to extract input from open source repositories for Athena. Current discussions are underway as how to support that effort for the

DOD. In addition, due to the new technologies for link extraction, SOCOM has asked to use AutoMap. They have asked us to support training there as well in fall 2009.

New extensions to AutoMap developed this year have been integrated into and taught in the dynamic network class at Carnegie Mellon University (CMU) and also at the Center for Computational Analysis of Social and Organizational Systems (CASOS) summer institute. Summer institute attendees included individuals from TRADOC and several intelligence agencies. They have also been taught to members of the National Guard, and were incorporated in a training module at the United States Military Academy (USMA). They were also taught to Canadian government personnel in Ottawa Canada.

Overview

Dynamic network analysis (DNA) is an emergent field centered on the collection, analysis, understanding, and prediction of dynamic relations (such as who talks to whom and who knows what) and the impact of such dynamics on individual and group behavior. DNA facilitates reasoning about real groups as complex dynamic systems that evolve over time. Within this field, computational techniques, such as machine learning and artificial intelligence, are combined with traditional graph and social network theory, and empirical research on human behavior, groups, organizations, and societies to develop and test tools and theories of relational enabled and constrained action.

A key aspect of DNA is the extraction of network data from texts. This involves the extraction of who, what, where, how, why, and when and the relations among given textual information that may contain a variety of typographical and content errors, translation mistakes, alternative wordings for the same event, aliases, etc. We propose to advance the state of text analysis, and admit the extraction of meta-networks from texts, and attributes of the extracted networks using DIME/PMESII cross classification. A combination of machine learning, rule based, and statistical algorithms will be used. In addition a combined semantic and syntactic approach will be used.

Research Approach

The application of DNA techniques to a large complex system, such as al Qaeda or the U.S. Army, entails a series of procedures. First, one needs to gather the relational data. One approach for doing this is to extract relations from a corpus of texts including open-source items such as Web pages, news articles, journal papers, stockholder reports, community rosters, and various forms of humint and signals intelligence (sigint). Second, the extracted networks need to be analyzed–that is, given the relational data, can we identify key actors and sub-groups, points of vulnerability, and so on? Third, given a set of vulnerabilities, we want to ask what would happen to the system were the vulnerabilities to be exploited. How might the networks change with and without strategic intervention? The CMU CASOS group has developed an interoperable suite of tools, shown in Figure 1, that acts as a chain to extract networks from texts, analyze these networks, and then engage in what-if reasoning.

This tool suite takes into account multi-mode, multi-link, and multi-time period data including attributes of nodes and edges. This toolset contains the following tools: AutoMap for extracting networks from texts, Organizational Risk Analyzer (ORA) for analyzing the extracted networks, and DyNet for what-if reasoning about the networks. This proposal is to improve the data extraction tool AutoMap and to adjust the analysis tool ORA so it can utilize the new types of data extractable via the improved AutoMap.
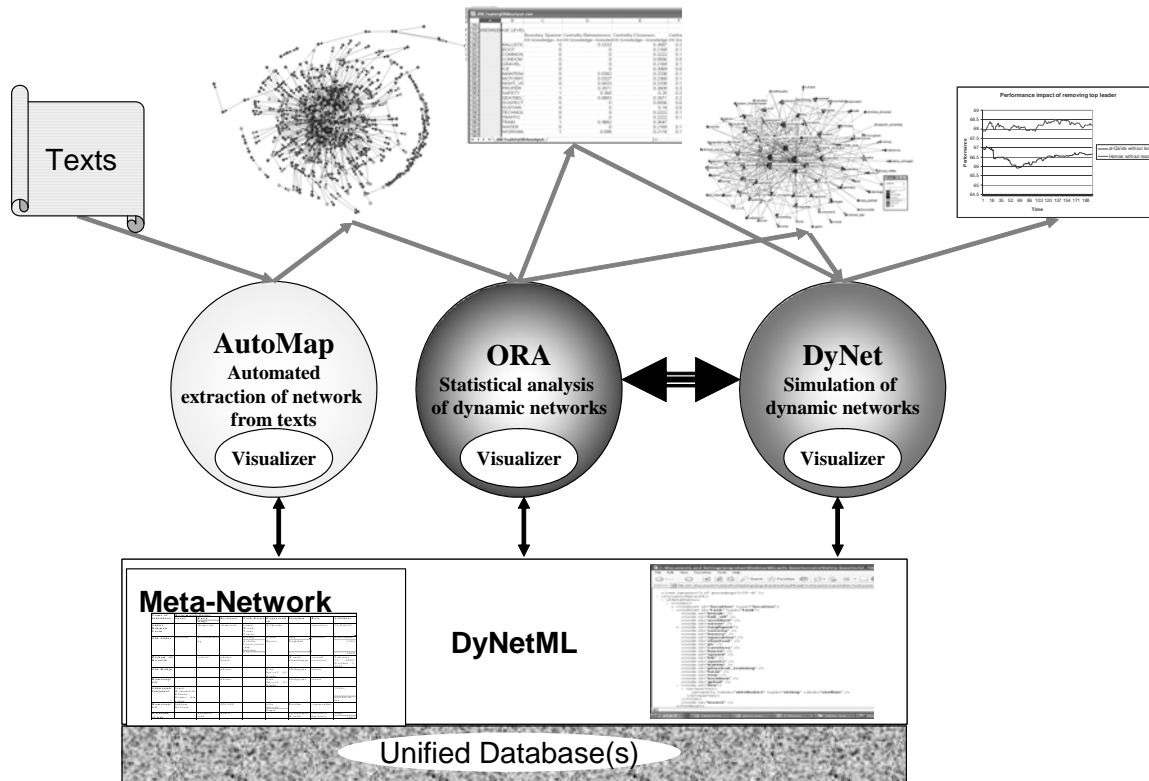

Figure 1. *CASOS tool chain.*

In terms of text-analysis, we expanded our approach this year. In the past we had focused exclusively on thesauri to cross-index terms and on proximity to define linkages. We focused on names entities for people, organizations and locations, and on categories for resources and knowledge areas. Anaphora resolution led to improved identification of such entities and the linkages among them. This year we augmented this approach with additional semantic information. We now also cross-tag terms by parts of speech and used a combined statistical and rule-based approach for cross-indexing terms and thesauri construction. We also utilized bi-gram identification for thesauri construction.

Accomplishments

Improved thesaurus construction in AutoMap; in particular, add new techniques for more automated thesaurus construction. We have developed a thesaurus editor to help users create and manage thesauri. Thesauri are constructed with an original text in mind, however thesauri are not often applied until after several preprocessing steps (stemming, generalization, etc.) are

performed. The thesauri editor provides tools to process the thesauri for use with preprocessed text, such as to include stemmed terms, ngrams, and words directly found in a concept list. Feedback also includes the identification of potential misspellings in the thesauri files.

Add tools to AutoMap to support alias detection, improved name-entity recognition, improved semantic reasoning. This will include but not be limited to, the use of parts of speech tagging and the use of machine learning techniques for cross-identifying words and their ontological categories. We have augmented our part of speech (POS) module for use with ill-formed text by creating and applying a look-up table procedure. Much writing uses improper grammar, spelling, and capitalization, which will lead to incorrect tagging when using a markov model approach to POS which has been trained on correct text. We have developed a system to take POS tagged text using our previous POS system to create a look-up table to apply POS to ill-formed text.

Add tools to AutoMap to extract information on time of activities discussed in texts and attributes of actors, events and locations. We have developed an interface to a gazateer database to extract location information and apply it as latitude and longitude information, which can be subsequently displayed geospatially by ORA.

Improve speed and scalability of AutoMap. Explore whether it is feasible to improve the speed at which AutoMap processes texts through multi-threading or parallelization. AutoMap has been refactored into a componentized form. This has resulted in a 10x speed-up in comparison to the previous version, mostly as a result of better file management. In the componentized form, AutoMap has significantly reduced its memory requirements. In addition, the script batch processing part of AutoMap has been augmented with an ability to specify a number of threads to run multiple batches on a multiprocessor system. The speedup is roughly proportional to the additional number of threads bounded by the number of processors on a machine.

Create merge and tracking routines for thesauri and databases constructed from different sources. AutoMap contains additional tools to aid in the management of supplemental data including a thesaurus editor and delete list editor as well as viewers for exploring the intermediate data produced by AutoMap. A "comparison" routine has been added to all of the editors/viewers to mark changes from previous versions including additions, removals, and also changes or modifications. For the editors this provides tracking routines from different sources, and from the viewers it provides insights into previous runs, especially those using different parameters.

Explore linking AutoMap directly to web-scraping technology. We have integrated a web scraper into AutoMap. The web scraper includes procedures to extract plain text from HTML files, PDF documents, Microsoft Word and Powerpoint files.

Explore linking AutoMap and ORA to streaming or regularly changing data streams. We have begun to develop a tool that will extract and continuously update data from streaming data sources such as web pages, blogs, emails, facebook, newsgroups, rss feeds, and wikis. The

import of streaming data will also be combined with our script batch system to allow for continuous update of data as well as the processing of data.

Develop processing routines to extract semantic information from meta-data, e.g., headers on email. Our tool to process streaming data (web pages, blogs, emails, facebook, newsgroups, rss feeds, and wikis) is being developed to extract both the unstructured text and also the structured headers. We are using the APIs of the data feeds to be able to ingest the semi-structured data rather than creating specialized parsers.

Augment DyNetML to handle meta-data information. The DyNetML specification has been extended to a type system closer to the user, such as continuous and categorical numbers, to more explicitly identify over time information. The AutoMap system has been modified to explicitly call the ORA back-end to enable faster adaptation of DyNetML advances.

Update help and add simple wizards to ORA and AutoMap to facilitate tool use by the military. AutoMap includes additional editors and viewers (explained earlier) that help to manage the external data files. An additional tool, the tagged text viewer, has been developed to interactively allow users to build generalization and metadata thesauri, and delete lists by clicking on words and adding them. Users are able to view the impact of delete and generalization upon the text. Metadata is color coded to provide visual feedback on how the concept terms are classified.
The icons in AutoMap have been reviewed to establish a consistent look for common items, such as a red x for delete list operations and files, a green arrow for generalization, and a purple swirl for meta-network operations and files.
An ORA wizard for data entry has been finished.
In addition a series of lesson plans for using AutoMap and ORA have been finished.

Develop and test procedures for co-clustering authors/speakers and what they are writing/talking about. This will enable identification of communities or practice. This has been finished and is in the new group-talk report. In addition, we extended the group-location report to handle better co-grouping.

Add tools to ORA so that while analyzing the data the user can link to the original data source, look at and augment the thesauri used for coding the data, and provide feedback to AutoMap. The DyNetML can contain pointers to text locations and then ORA can display the text. AutoMap has been adjusted so that it puts the pointer in DyNetML automatically. The user can in ORA in the visualize click a node or link and then see where the data comes from.

Test tools by analyzing data collected from: We have done extensive testing this year. In addition to these sources used last year,
- The Web for information on various terrorist groups,
- Enron corpus for email,
We have also tested the system with 43,000 texts on the Sudan, and various data on the pacific rim. We just received a set of data from OMNI-FUSION and will test the system on that next year. That will be an interesting challenge as that is chat and email data, rather than news articles and Enron email.

Explore and develop emotion-based analysis of semantic networks and communicative power of key concepts. We identified emotion based lexicons for data concept analysis..

Provide the tools and data from the battle lab regularly to West Point for use by cadets under Lieutenant Colonel John Graham. We have provided all tools to the USMA and they are being used. We also have sent tools periodically to various people at ARI and ARL, and to SOCOM.

Utilize feedback from cadets, intelligence analysts, and others to improve user-interface. This is ongoing.

In addition to the basic work we have also built an initial sentiment extractor. We are currently exploring the use of various web crawling techniques to build our own bigram list for world leaders, locations etc. to support military coding and automated updating of thesauri.

Contributions to Basic Science

Interest in social network science is continuing its decades-long rise mainly due to its usefulness in an assortment of broader research disciplines, such as organization behavior, sociology, computer science, and forensic science. A key problem is that network science tends to be applied primarily to data collected via surveys or from membership manifests. This means that vast quantities of ethnographic and qualitative data go unanalyzed, as they can currently be processed only through time-consuming and error-prone manual coding. The tools proposed will make it possible to assess and analyze more of the information in these texts, reduce coding barriers, and facilitate rapid automated and semi-automated coding.

The specific work this year improves the accuracy of link extraction through anaphora resolution and parts of speech processing. This is critical as without this, estimations of connections among individuals and groups are grossly over-stated. The refactoring we did is critical as it enables the system to operate in parallel and to work much faster. This means that we can now process 10's of thousands of texts in a few hours, where it used to take weeks. The bi-gram work and parts of speech processing are now enabling more rapid meta-network thesauri construction. This means that we can often classify concepts as whether they represent a who, what, how, why, where without human intervention. This reduces coding time and increases accuracy.

Potential Army/Military Applications

The initial applications of these tools for the military are in human terrain assessment/ cultural geography assessment, rapid identification of issues and sentiments in new areas, and change in attitudes in at-risk populations. Additional key military applications include: assessment of hot topics under discussion in foreign countries, political and social attitude assessment, identification of key opinion leaders and changes in their opinions, coding of reports by commands and leaving them as reports to future commands in an extensible and usable form, assessment of degree of misinterpretation of commanders' intent, assessment and cross-classification of rules of engagement. In addition, this year, discussions with Ben Jordan and

others suggest that it should be possible to expand AutoMap through event identification and system indicators to build systems for auto-identifying from texts the input values and parameters needed for many human socio-cultural simulation models such as Athena.

Future Plans

We will continue to refine and improve on current activities, described in the accomplishments section, including:

- Test tools by analyzing new large data corpi. If you have data you want us to process, please let us know. We would be happy to jointly collaborate on a project. You could, e.g., send us electronic versions of all reports and we could assess them with these tools.
- Utilize feedback from SOCOM, TRADOC, cadets, intelligence analysts, and others to improve user-interface.
- Improve sentiment extraction, event extraction, and temporal indicators extraction.
- Continue to explore linking AutoMap and ORA to streaming or regularly changing data streams.
- Improve speed and scalability of AutoMap further through the use of a high throughput computing, such as is done with a distributed collection of computers.
- Build emotion-based valence indicator of messages.

In addition, we will begin to focus on the following new tasks:

- Syntactic parsing to improve semi-automated thesauri construction
- Explore machine-learning techniques for improved auto-theme extraction.

Virtual Worlds: An Exploratorium for Theorizing and Modeling the Dynamics of Group Behavior

Contract #: W91WAW-08-C-0106   **Contract Dates:** 9/29/2008 to 9/28/2010
Institution: Northwestern University   **PI:** Noshir Contractor
Co-PIs: Marshall Scott Poole, Dmitri Williams, Jaideep Srivastava

Problem(s)/Research Question(s) – Can we specify the precise nature of relationships between various factors such as group composition and group experience and group performance? How do variables such as task complexity and environmental stress influence these relationships?

Technical Barrier(s) – While much research has been conducted on factors that influence group performance, such research has typically gathered only a few data points. For example, an experiment might have two levels of group composition or experience, or two or three levels of complexity. In studies that gather information on real world teams, there are typically less than 100 teams. These limit our ability to determine the shape of the curve that relates composition and performance or composition, task complexity, and performance.

Significance/Impact for Basic Research – Current knowledge about the impacts of various factors on group performance is sketchy and does not enable us to specify relationships precisely. Deriving more precise response surfaces of the joint relationships among composition, experience, complexity, and stress with performance would greatly enhance our understanding of groups and teams. It would enable us to evaluate theories of team effectiveness based on less than complete data and to respecify relationships based on complete evidence.

Potential Transitions Knowledge gained in this basic research may be usefully employed in ARI's applied training research programs (e.g., Future Force Warrior Training, Managing Training and Providing Feedback within Networked Embedded Training)

Overview

This series of quasi experiments will derive detailed curves of the relationship between group composition and experience and group performance under varying conditions of task complexity and stress. It will draw on samples of two to three thousand groups varying in composition and experience level in the game. They are engaged in a variety of tasks that vary in complexity and amount of stress the team is under. Rather than having a few points along the continuum of composition, experience, task complexity, and stress, the sample will allow us to have a wide range of values of each. This will enable us to determine if relationships are linear, curvilinear, U-shaped, or stepladder type, among other possibilities. An illustration of a hypothetical response surface is shown below.
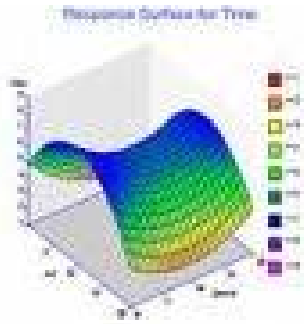
This picture of a hypothetical response surface shows a response variable (which in our case will be group performance) plotted as a function of two independent variables. If we can successfully derive these response surfaces for group performance, it will give us clues as to how performance will respond to varying levels of factors such as experience, task complexity, etc.

Research Approach

The basic approach is to identify groups in a database downloaded from EverQuest2, a massive multiplayer online game (MMOG), and then code them for: (a) composition in terms of expertise and role in the game; (b) level of common experience with members of the groups they join; (c) complexity of the task they are engaged in; (d) stress level of the task; and (e) group effectiveness, as measured by whether they successfully accomplish the task and the level of rewards they succeed in gaining due to accomplishment.of the task. We will then conduct two types of analysis: (1) We will test structural equation models of the causal impacts of the various factors on performance. This will enable us to test and validate various models of teamwork effectiveness that have been previously posited. (2) We will develop response surface models of the relationships between combinations of the factors and team performance. This will provide insight into the specific nature of the relationships among the factors.

Identification of Groups. This has taken some time because the SONY database contains information useful to system administrators and is not explicitly designed for research. So we have had to induce groups based on the information SONY chose to preserve. We are employing two approaches to identifying groups: (1) find a set of players who finish a task and receive a reward at the same time and place and (2) find a set of players who are consistently together in the same places at the same time for given period of time (the first set is a subset of the second because groups do not always succeed at their tasks, nor are they always engaged in a goal-directed task). Identifying groups based on the first method is fairly straightforward, but it yields only those groups that succeeded. This is an important set of groups to sample. However, we are also pursuing the second method, because it promises to yield a much larger set of groups including both those that succeed at their task, but those that fail, and those engaged in other activities. The limitation of the second strategy is that it yields "false positives," sets of players who happen to be in the same place, but do not really belong to the same group. We are developing strategies to deal with this.

Currently we can easily draw samples using method 1 and we have conducted two investigations using this method. We have been working for 4 months on developing systems to identify groups using method 2.

Expertise. Our first composition variable is expertise in the group. Much research suggests that teams composed of highly expert members outperform those with lower levels of expertise. In an ideal world, we could compose teams of the best experts available. However, in the real world, we must often build teams from the available personnel, who may differ widely in expertise. This raises numerous questions such as the following, which we will illustrate with for a four-person group:

- What proportion of the group must be highly expert before performance degrades significantly compared to a group comprised of experts? Is a group with two experts as effective as one with four? Is a group of four moderate experts better than a group with two high experts and two low in expertise?
- How does task influence the impact of expertise on performance? For high complexity tasks does performance degrade more rapidly for groups with fewer experts than for low complexity tasks? If experts hold the critically important roles, is performance less likely to degrade when less expert members hold the other roles?

In order to pursue these and other questions, our measure of expertise has two components, as defined by Sonnentag (2000): (1) *Accomplishment* as measured by an index composed of the player's statistics on several important attributes that contribute to ability to contribute, including strength, wisdom, agility, intelligence, and stamina. Players build these up as they progress in the game and they reflect the experience component of expertise defined by Sonnentag. (2) *Performance ability* as measured by the speed with which the player moves up in levels compared to the average speed of moving up in levels. This reflects the ability and skill components of expertise defined by Sonnentag.

Indices of expert composition include average level of expertise in the group, total expertise in the group, and distributions of experts based on splitting the data.

Experience. Experience in the group is the second group composition variable. While this can be defined as a team level variable, it is also possible to define it in terms of composition of members with particular levels of expertise with each other, because members can switch teams easily in EverQuest 2. In this respect EQ2 groups are similar to airline crews or surgical teams, which are composed of roles (e.g. surgeon or nurse) whose incumbents are not always the same people. Hence the same group may be composed over time by different people with different level sof experience with each other.

Experience will be measured in three ways. For groups identified using method 1 as described above, experience is measured in terms of how long they played together in a given session. For groups identified using method 2 described above, experience is measured in terms of how long they have met with their fellow members since the beginning of the period covered by the data to the time when the group is working on its task. A third measure of experience is the total amount of time each dyad that makes up the group has played together since the beginning of the period covered by the data to the time when the group is working on its task.

Task Complexity. This will be measured by having coders scan sites associated with EverQuest2 for information about the various quests, the type of task this research will focus on. This information, which takes qualitative and quantitative forms, will be used to classify tasks in

terms of complexity in terms of the number of different elements the task has, how difficult each of these elements is, how tightly elements interact, and how dynamically the elements change. The greater on all these four dimensions, the more complex the task.

Stress. This will be measured in terms of the discrepancy between the levels the members of the group have attained and the level the task is designed for. In EQ2 players must be nearly at the level of their task for them to be able to take it on. A task in which the members' average level is at one or two levels above the level of the task will be classed as low stress; when the average group level equals the level of the task, it will be classed as moderate stress; when average level is one below the level of the task it will be classed as high stress; and when group's average level is two below the level of the task it will be classed as very high stress.

Accomplishments

We have conducted two tests of structural equation models of effects of several of the factors on effectiveness with groups sampled using method 1 (Xiong, Poole, Williams, & Ahmad, 2009; Zhu, 2009). These have shown that length of time playing together, frequency of playing together, higher levels of expertise, and complementary roles on the team is positively related to effectiveness. On the other hand, communication is negatively related to effectiveness and playing with multiple other players is not correlated with group effectiveness. While perhaps only the last of these findings is a particularly novel, they replicate expected relationships in real world groups. This provides evidence that the groups in EQ2 have at least a degree of verisimilitude to real world groups. Future studies will try to tease out in more detail the nature of these relationships in terms of response surfaces.

We have developed algorithms for capturing groups using the second method of group definition. These have indicated that there are a wide variety of types of groups in EQ2 based on interaction history. Groups often work on their own, then combine with other groups in larger groups, then break down. Tracking the data longitudinally offers us opportunities to take the groups' interaction context into account. We have drawn a large sample of several hundred groups working together over a week's play to test some preliminary performance models. These results are expected during the first part of 2010.

Contributions to Basic Research

As previously noted, the large samples of action teams that we can draw will enable us to evaluate models of group effectiveness with much greater certainty than research with smaller samples of groups. It will enable us to estimate the relative contributions of different factors with more confidence.

The response curve analyses should enable us to discover new facts about group performance, including whether relationships are linear, curvilinear, or have sudden thresholds. These facts should both aid in evaluation of current theories/models and suggest new ones.

Potential Army/Military Applications

The results have two types of implications for Army-Military applications. First, they can guide team composition. Currently the composition of military units is more of an art than a science. Soldiers are assigned and trained for their assignments and it is presumed that this is all that can be done to ensure team effectiveness. Provided valid and reliable measures of expertise and experience for military personnel can be derived, the results of our research can be used to guide applied research on the assignment of personnel to units. While constraints of logistics and availability of appropriate personnel will always place a limit on optimal assignment, this research will offer some guidance when there is room for maneuver.

Second, this research can lend some advice to those planning and managing collective military training. It should yield some insights as to proper combination of personnel during training. For instance, if it is better to train with the same unit members than to shift training to different combinations of personnel, this has clear implications for training design. If the opposite conclusion is supported, then training must be designed in fundamentally different ways.

Future Plans

We will develop formal mathematical and simulation models that may lead to applied research validating our models of personnel assignments to training and to units. In addition, we hope to be able to develop contingency models to suggest which strategies for performance are most appropriate under varying conditions of task complexity and stress.

References

Sonnentag, S. (2000). Expertise at work: Experience and excellent performance. In C. I. Cooper & I. T. Robertson (Eds.) *International review of industrial and organizational psychology* (vol. 15, pp. 223-264). London: John Wiley & Sons.

Xiong, L., Poole, M. S., Williams, D., & Ahmad, M. (2009, July). *The effects of group structure on group behavior and outcomes in an online gaming environment.* Presented at the INGRoup Conference, Colorado Springs, CO.

Zhu, M. (2009, October). *Assembly of successful teams: Insights from the study of massively multiplayer online role playing games (mmorpgs).* Presented at the Midwest Organizational Communication Miniconference, Evanston, IL.

Enhancing Employee Dedication and Retention: The Contribution of Perceived Organizational Support

Contract #: W91WAW-08-C-0031
Institution: University of Delaware

Contract Dates: 02/15/2008 to 02/14/2011
PI: Robert Eisenberger
Co-PI: Louis Buffardi, George Mason Univ.

Problem(s)/Research Question(s) – Our project seeks to address three related questions:

- How do recruitment promises contribute to new employees' perceptions of being valued by the organization (perceived organizational support, or POS) and information seeking?
- Does employee POS overcome negative employee reactions that occur when the organization cannot fulfill all its promises?
- How can supportive treatment and affective commitment that employees associate with lower level organizational units be used to enhance commitment to higher level units?
- Will a practical intervention that instructs supervisors on providing support for subordinates and encouraging subordinates to view them as symbolic of the organization increase the subordinates' performance and job retention?

Technical Barrier(s) – The organizational support theory approach to employee-employer relationships has only recently been extended to the research issues.

Significance/Impact for Basic Research – The present research will produce a greater understanding of the influence of perceived organizational support on employees' information seeking and reactions to unmet promises, including consequences for performance and turnover. Our findings will also bear on the issue of how employees' commitment to supervisors and work groups generalizes to higher level organizational units (departments and the overall organization).

Potential Transitions – Knowledge gained in this basic research may potentially be used in:

- ARI's applied research program on Strategies to Retain Quality Enlisted Soldiers and Officer Accession and Retention (Applied Broad Agency Announcement 3-B and 3-C).
- ARI's applied research program on Tomorrow's Leaders, Leading the Future Force (Applied Broad Agency Announcement 2-A).

Overview

Organizational support theory (Eisenberger et al., 1986) holds that employees acquire a general perception concerning the extent to which the organization values their contribution and cares about their well-being (perceived organizational support, or POS). Employees use POS to meet socio-emotional needs (e.g., approval, esteem, and emotional support) and to gauge the organization's readiness to reward greater effort (Eisenberger et al., 1986). Promises made to prospective employees should increase POS (Aselage & Eisenberger, 2003). An important positive consequence of heightened POS might be proactive information seeking wherein

employees take an active role in learning more about their jobs and adjusting to the organization (Wanberg & Kammeyer-Mueller, 2000). Thus, we hypothesize:

- H1: Pre-entry promises and traditional favorable practices by an organization will enhance employees' perceptions that the organization has imposed obligations on itself to treat employees' favorably, leading to greater POS which should, in turn, increase proactive information seeking, affective organizational commitment, citizenship behaviors, and retention.

Another major theory of employee-organization relationships (psychological contract theory) emphasizes the aversive consequences of the organization's failure to live up to obligations perceived by employees as they enter the organization (Rousseau, 1995). According to psychological contract theory (Rousseau, 1995), during recruitment and early phases of employment, employees develop a set of understandings concerning the organization's obligations. Violations of this *psychological contract* have been found to be associated with negative emotionality, reduced organizational citizenship behavior, and increased voluntary turnover (Robinson & Morrison, 2000; Rousseau, 1995). Recent research suggests that POS can mitigate the negative effects of the organization's failure to carry out specific promises among current employees (Coyle-Shapiro & Conway, 2004). POS might play an important role in maintaining employee commitment and retention when the organization is unable to keep some of the promises made during recruitment.

- H2: High POS will reduce the negative influence of unmet pre-entry promises on affective organizational commitment, citizenship behaviors, and retention.

Organizational support theory sheds light on the means by which favorable treatment received from supervisors and lower level units is credited to higher level units, including the organization as a whole. Stinglhamber, Eisenberger, Aselage, Jones, Sucharski, Becker, and Eder (2007) suggested that organizational units (e.g., supervisors, work groups, departments) differ systematically in the extent to which they are perceived by employees as synonymous with the organization. This concept of organizational embodiment has important implications for organizations because organizational embodiment may influence the extent to which favorable or unfavorable treatment by any nested organizational unit (e.g., supervisor, work unit) affects organization members' POS and affective commitment associated with higher-level organizational units (e.g., department, entire organization).

- H3: High organizational embodiment will increase the generalization of perceived support and organizational commitment from lower order organizational units (supervisor and work group) to higher order organizational units (department and organization).

Finally, we propose an intervention designed to increase supervisors' organizational embodiment and supportive behaviors, which should enhance subordinates' POS and affective commitment to the organization and decrease turnover. We will teach supervisors more effective ways to provide support for subordinates through mentoring, feedback, and praise for superior performance. We will also teach supervisors to increase subordinates' perceptions of them as synonymous with the organization.

- H4: The subordinates of supervisors who receive the training and who have positive beliefs in the organization's supportive orientation toward employees will show the

greatest increases in perceived supervisor support and POS, and the greatest affective commitment, citizenship behavior, and voluntary retention.

Research Approach

We are conducting three longitudinal field studies in civilian populations that include graduating seniors accepting full-time jobs, employees in a hierarchical organization, and a social service agency. Surveys will be administered over time to assess the relevant attitudinal variables. Retention data will be taken from organizational records. The research approach, based on prior empirical findings, extends the tenets of organizational support theory in order to address the new set of questions.

Accomplishments

Investigation 1, concerning the work experiences of college graduates during their first year of employment, was begun in the Spring of 2008. We received back 212 completed questionnaires by students who had received job offers but had not yet started work (Wave 1), 169 competed questionnaire in Wave 2 (3 weeks following the beginning of employment), 147 in Wave 3 (3 months following the beginning of employment) and 128 in Wave 4 (6 months following the beginning of employment). In order to have sufficient numbers respondents for statistical analyses we have carried over the data collection into the current academic year with an additional sample of college graduates. Thus far we have received 192 completed surveys in Wave 1 and 119 completed surveys in Wave 2, with Wave 3 underway. These numbers should be sufficient for our statistical analysis when data collection is completed in December.

Investigation 2 involves increasing our understanding of the generalization of employees' perceived organizational support from lower order organizational units (supervisor and work group to higher organizational units (department and organization). The first wave of this two-wave longitudinal data collection has been completed with 286 employees of a social services agency. Initial analyses of the first wave of data indicate that consistent with H3, employees' identification of their supervisors with the organization (supervisor's organizational embodiment) plays an important role in employees' extension of favorable relationships with supervisors to the organization (Eisenberger, R., Karagonlar, G., Stinglhamber, F., Neves, P., Becker, T. E., Gonzalez-Morales, M. G., & Steiger-Mueller, M. [In press]. Leader-member exchange and affective organizational commitment: The contribution of supervisor's organizational embodiment. *Journal of Applied Psychology*.)

Investigation 3 concerns an intervention to increase supportive behaviors of employees and the extent to which the supervisors identify their favorable actions with the organization, which should enhance subordinates' perceived organizational support and affective commitment to the organization and decrease turnover. We have gathered critical incidents of support and nonsupport from employees of a chain of restaurants. We are using this information to help develop the protocol for teaching supervisors to treat subordinates more supportively in preparation for intervention in which half the restaurants will receive the intervention and the remainder will receive no treatment and serve as a control condition.

Contributions to Basic Science

We anticipate that the present research will produce a greater understanding of the influence of perceived organizational support on employees' information seeking and reactions to unmet promises, including consequences for performance and turnover. Our findings will also bear on the issue of how employees' commitment to supervisors and work groups generalizes to higher-level organizational units (departments and the overall organization).

Potential Army/Military Application

Organizational support theory, which emphasizes organization members' perceptions of their valuation by the organization, provides a basis for understanding major influences on Soldiers' citizenship behaviors (voluntary actions aimed at helping the organization reach its objectives) and retention. The Army faces the difficult problem of increasing recruitment, citizenship behavior, and retention at a time when Soldiers' careers often require long periods away from family, prolonged exposure to stressful environments, and short recuperation periods. Further, based on policy decisions that are often beyond the military's control, high demands on personnel may result in training opportunities and assignments that compare unfavorably with what has been explicitly promised or expected based on the Army's traditions. Organizational support theory suggests that Soldiers who feel supported will more likely seek out needed job related information, perform better, be less upset by unmet promises, and be more likely to re-enlist. Our research will provide evidence with employees on these predictions. If our hypothesized findings are confirmed, procedures can be developed to train officers to enhance perceived organizational support in Soldiers that will increase performance and retention.

Additionally, the generalization of Soldiers favorable orientation from the local organizational unit (e.g., squad or platoon) to higher-order organizational units (e.g., division and the Army itself) is a key issue for the Army. An effective Army leader of a squad, platoon, or regiment engenders in subordinates an emotional bond to himself or herself and the unit. Because different units must cooperate on common missions and because Soldier retention is an essential Army goal, it is important to understand how leadership and human resource practices that contribute to affective commitment and performance at the local level affect a favorable orientation toward higher-level units. Our findings will provide a basis for the development of methods to train officers to increase the generalization of positive regard and commitment from lower level Army units to the Army as a whole.

Future Plans

In the coming year, we anticipate finishing data collection and analysis of our longitudinal study concerning the first two research issues. The results will be written as an ARI report, prepared for presentation at a professional conference, and preparation begun for an article in a scientific journal. Work will also be carried out addressing the third research issue involving the generalization of organizational commitment from lower order organizational units to higher order units. We will develop appropriate measures for each scale, and begin administering surveys including these measures.

Successful Collaboration at a Distance

Contract #: W91WAW-07-C-0060　　　　**Contract Dates:** 08/13/2007 to 08/12/2010
Institution: University of Michigan (UC Irvine)　**PI:** Gary M. Olson

Problem(s)/Research Question(s) – While modern communication and computing technology make it possible for groups or teams to work together when their members are in different locations, it is by now well known that it is difficult to do so. A variety of factors have been identified as possible reasons for this, but research is needed to clarify the role of these factors, including the relative importance of each.

Technical Barrier(s) – Identifying why it is difficult to work at a distance has required the collection of large amounts of data in both the field and the laboratory. We have developed a set of working hypotheses about what factors affect the success or failure of long-distance collaborations. While the total set of factors is quite extensive, they fall into five broad categories: Collaboration Readiness, Technology Readiness, Common Ground, Management, and Nature of the Work. Specifics about the full set of factors can be found in Olson, et al. (2008).

Significance/Impact for Basic Research – Many researchers are interested in this problem, but to date, no comprehensive account exists.

Potential Transitions – We hope to provide those with goals of carrying out successful collaborations at a distance with specific guidelines for how to maximize the chances of success.

Overview

This research seeks to understand the factors that contribute to success or failure in geographically distributed projects. In earlier work, we generated a series of working hypotheses about factors that are important in geographically distributed projects. In this project, we seek to verify these working hypotheses and determine their relative importance.

Research Approach

We are carrying out four lines of research in this project:

1. Research efforts in the field of teams working at a distance.
2. Laboratory experiments of selected factors,
3. Computational models of the task involved in the laboratory experiments, to help select which variables to study empirically, and
4. The construction of a collaboratory online wizard that will be a resource for those about to embark on a geographically distributed collaboration, to assess the risks involved in their project and receive recommendations as to what to do about those risks.

Accomplishments

We have moved this project to the University of California, Irvine, though the funding continues through the University of Michigan, with a subcontract to us at UCI. Our progress to date is as follows (and corresponds to the numbered items under Research Approach above).

1. We have carried out interviews with teams in companies that are attempting to work together at a distance. We have gathered 120 interviews from 16 sites so far. We also obtained a rich set of interviews from Avaya, who themselves had conducted similar interviews for a large-scale project of theirs. We are writing papers that use these data, and have committed to a monograph with Morgan & Claypool that will be completed this summer.

2. We completed a study at the University of Michigan, and a paper reporting this work is currently under review. With start-up funds provided by the University of California, Irvine, we have built a new lab at UCI that is now operational. We have a new study that we are pilot testing in the new lab, and will begin running the next experiment in this series shortly.

3. The computational modeling is still a future project, awaiting the recruitment of a student with the appropriate computational skills.

4. We have made major progress on what we are now calling the Collaboration Success Wizard. We moved all the software to UCI, reprogrammed much of it, and carried out user testing on it. In the summer of 2009 we did a major pilot test of the Wizard, administering it to 43 members of a proposed project in the UCI Medical School. We analyzed the data and provided the project with feedback about their prospects, which they then incorporated into their large proposal to NIH. We are about to embark on a second major test of the Wizard with a consortium of collaborators headquartered at the University of Pittsburgh. A major public announcement of the Wizard will be made in late April at a workshop on the Science of Team Science hosted by Northwestern University. These initial efforts with the Wizard will be included in the monograph mentioned in item 1 above. We expect to release the Wizard to the public within the next month or two.

Contributions to Basic Science

Given how much interest there is in questions of the social and technical barriers to successful collaboration at a distance, the work in progress will have far ranging implications for a number of science areas. We expect the Collaboration Success Wizard to be a major source of data for our efforts.

Potential Army/Military Applications

Our findings should be of great value to the Army and the military since geographically distributed teamwork has emerged as a major thrust across all sectors. In particular, the

Collaboration Success Wizard should be a useful tool for Army projects.

Future Plans

We are continuing work in all four areas, and expect substantial accomplishments in all of them now that the project is up to full steam at UC Irvine. We have also just received major funding from Google to extend this work to cultural aspects of collaboration, which will draw on the insights achieved so far in this current project.

References

Olson, J. S., Hofer, E. C., Bos, N., Zimmerman, A. Olson, G. M., Cooney, D., et al. (2008). A theory of remote scientific collaboration (TORSC). In G. M. Olson, A. Zimmerman, & N. Bos (Eds.) *Scientific collaboration on the Internet.* Cambridge, MA: MIT Press.